AcaStat User Manual

Version 6.1

© Copyright 2009, AcaStat Software.
All rights Reserved.
http://www.acastat.com

Table of Contents

INTRODUCTION 5

PURPOSE 5
GETTING HELP 5
ACADEMIC APPLICATION 5
INSTALLATION 5
UNINSTALLING ACASTAT 6
ACASTAT STARTUP SCREEN 6
UPDATES 7
CALCULATOR 7
OPTIONS PANEL 8

DATA GRID 9

Rows 9
Columns 9
Cells 10
Entering data 10
Editing data 10
Right-Click Support 10
Spreadsheet and Data File Limits 10
Split Data Module 12
Importing Data 13
Pasting Table Data 13
Copying data from Microsoft Access 14
Importing Text Data Files 14
Tips on using a spreadsheet application to create a csv or txt file 15
Example of Import Error 15
Exporting Data 15
Sending Data to Excel 15
Saving as a Text File 16
SYLK Errors 16
Changing Variable Names 16
Find Cell Contents 16
Find and Replace 16
Adding/Deleting Rows and Columns 17
Add/Delete Rows 17
Add/Delete Columns 17
Using a Weight Variable 17
Formatting Variables 18
Variable Name 19
Variable Label 19
Change Decimals 19
Value Labels 19
Missing Values 19
Recoding Values 19
Date Values as Controls 19
Computing New Variables 20
Comparing Dates 22

Combining Two Variables 23
Trimming Extreme Values 24

STATISTICAL PROCEDURES 25

Variable Control Panel 25
Dependent Variable 25
Independent Variable 25
Control Variable 25
Creating Data Subset with Controls 26
Statistical Procedures 26
Frequencies 26
Descriptives 26
Explore 26
Crosstabulation 26
Pearson Correlation 27
Regression 27
List Variables 27
Diagnostic Accuracy 27
T-tests of means 28
Analysis of Variance (ANOVA) 28
Summary Report 28
Simulation 31
Random Sample 31
Repeated Random Sampling 31
Appraisal Ratios 31
Project Notes 32
Data File Information 32

OUTPUT 33

Page Breaks 34
Printing 34
Exporting 34
Copying and Pasting Output 34
Converting Output to Tab Delimited 34

CHARTS 35

SUMSTATS 36

SUMSTATS TIPS 36
t-test Module 37
One Sample Mean 37
Two Sample Means 37
ANOVA Module 37
Z-test Module 37
One Sample Proportion 37
Two Sample Proportions 37
Chi-square Module 37
Margin of Error for Means Module 38
Margin of Error for Proportions Module 38
Standardized Score Module 38
Diagnostic Accuracy Module 38

Power Module 39
GLOSSARY 40
How Glossary Works 40
Glossary File System 41
Searching Records 41
Limitations 41
File Menu Items 41
New File 41
Open File 41
Save File As 41
Add Records from Other Files 42
Import 42
Export 42
Print 42
Edit Menu Items 42
Add Note 42
Save Note 42
Delete Note 43
Copy 43
Paste 43
Search 43
Reset Search 43
PRACTICE DATA FILES 44
Repair Practice Data Files 44
Practice (csv) Delimited Data Files 44
MANUALS (PDF) 45
User Manual 45
Student Workbook 45
Applied Statistics Handbook 45
REVIEW OF FEATURES 46
Data Grid 46
SumStats 47
Output Log 47
Glossary 47
Handbook 47
SYSTEM REQUIREMENTS 48
PURCHASE INFORMATION 49

Introduction

Purpose

AcaStat provides students, teachers, researchers, and managers with an inexpensive and easy to use data analysis tool and instructional aid. AcaStat combines several modules into one convenient package. These modules include Data Grid, Charts, SumStats, Glossary, and the Research Methods Handbook.

Getting Help

Both application and statistical help are provided throughout the AcaStat modules. Click on any help button to get help on the current module.

In addition to application help, statistical help is provided in three formats:

- ✓ The Research Methods Handbook is available in html format in the Handbook module.
- ✓ A glossary is provided for statistical terms and definitions. You can add to this glossary or create your own note database.
- ✓ An Applied Statistics Handbook is provided in pdf format for Licensed Edition (LE) versions.

Academic Application

For those interested in using AcaStat as part of their academic teaching/learning, an excellent companion text would be Hubert M. Blalock Jr.'s text Social Statistics, 2nd Edition (McGraw-Hill Publishing Company). Most of the formulas used in the statistics Handbook are consistent with Blalock. Practice problems are provided for most of the statistical methods contained in the Handbook. When conducting manual calculations, students should round each calculation to three decimal places (e.g., 1.5865 should be rounded to 1.587) to obtain results consistent with the Handbook.

Installation

Download AcaStat from http://www.acastat.com. The software is downloaded in a self-extracting file. Save the software to your hard drive and double-click on the file to begin the installation process. Some installations require a 1.5 megabyte download of installation software. This will be done automatically during the installation and will require an internet connection. Software purchased on a CD does not require an internet connection.

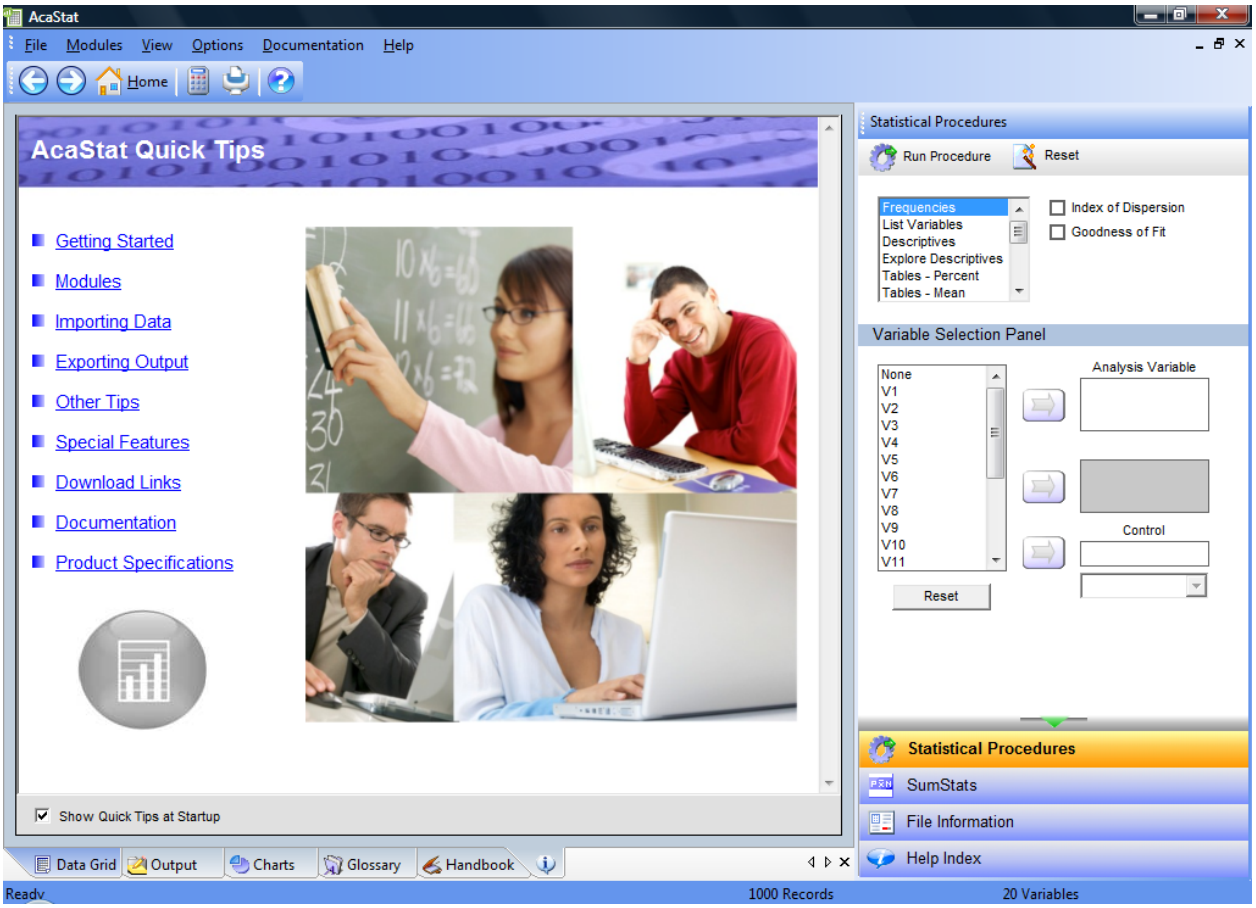
Uninstalling AcaStat

Use the Windows Control Panel option ADD/REMOVE programs to uninstall this software. Files that you have created or changed cannot be removed by the ADD/REMOVE option. These files can be easily removed by manually deleting them from the hard drive by using Explorer or My Computer to highlight and delete the 'AcaStat' program folder in the 'C:\Program Files' subdirectory.

AcaStat Startup Screen

AcaStat starts by displaying the Quick Tips tab. There are also tabs for Data Grid, Output, Charts, Glossary, and Handbook. In addition, a control panel contains Statistical Procedures options, SumStats, File Information, and Help Index. Use the Statistical Procedures panel to select and run summary statistics, basic descriptive and inferential statistics, and crosstabulations. The Output tab will automatically display the results for the current analysis or a log of all analyses since you opened AcaStat. The output from all procedures except Summary Reports will be automatically appended to the Output Log.

The Option pull-down menu allows you to hide all tabs except Output. These tabs will automatically reappear when you restart AcaStat.



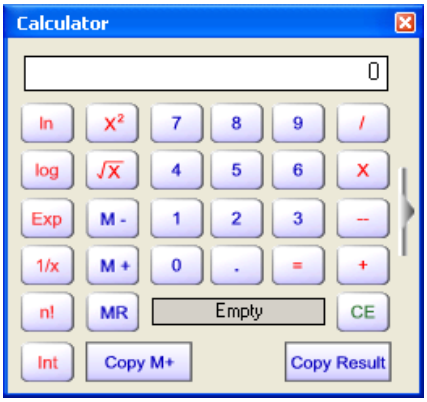
Updates

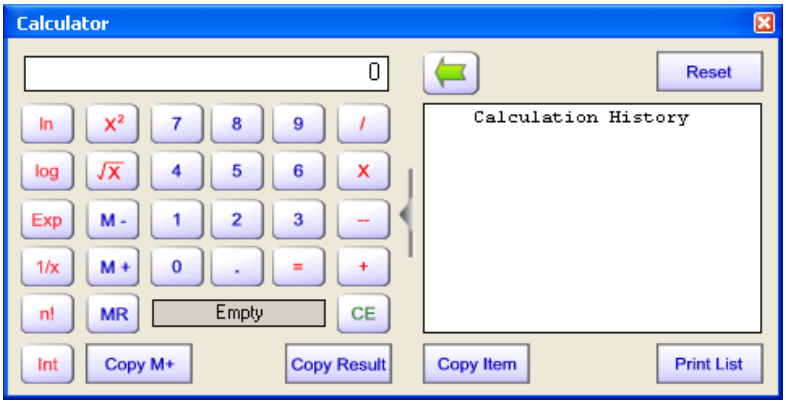
The Help menu contains an option to check for updates. AcaStat will connect with the Internet and automatically compare your installed version with the version of the update, if any. If a newer update is available, the appropriate files are downloaded to your computer.

Update - AcaStat
Welcome to the AcaStat Update Wizard
This wizard will check available updates for AcaStat.
Click Next to continue, or Cancel to exit update.
Click here to configure the proxy
< Back
Next >
Cancel

Calculator

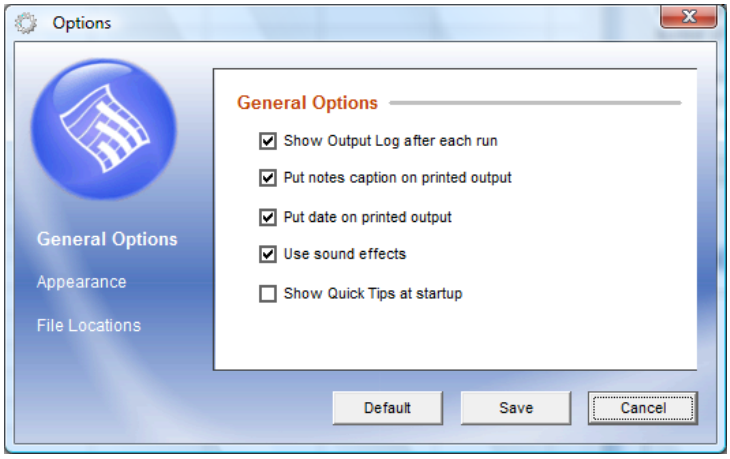
A pop-up calculator is available on the toolbar. It is a stay-on-top design that remains in view when using AcaStat. The calculator allows copying of data in the result field or memory storage field.

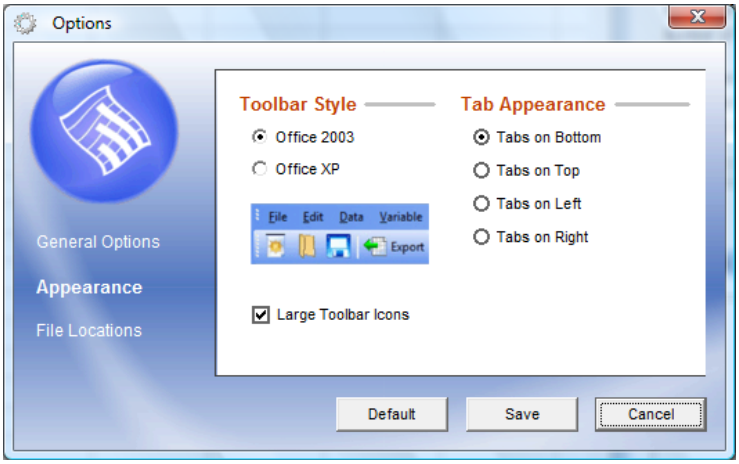


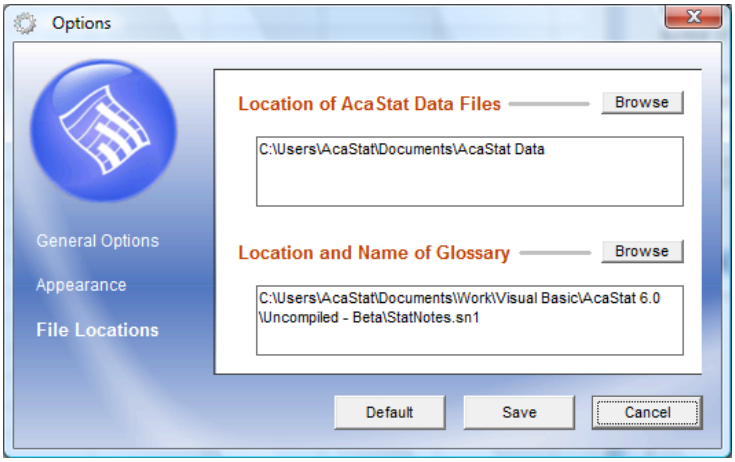


Options Panel

The options panel allows you to change output settings and the location of data files and glossary file. Some changes may not be complete until you close and reopen AcaStat.

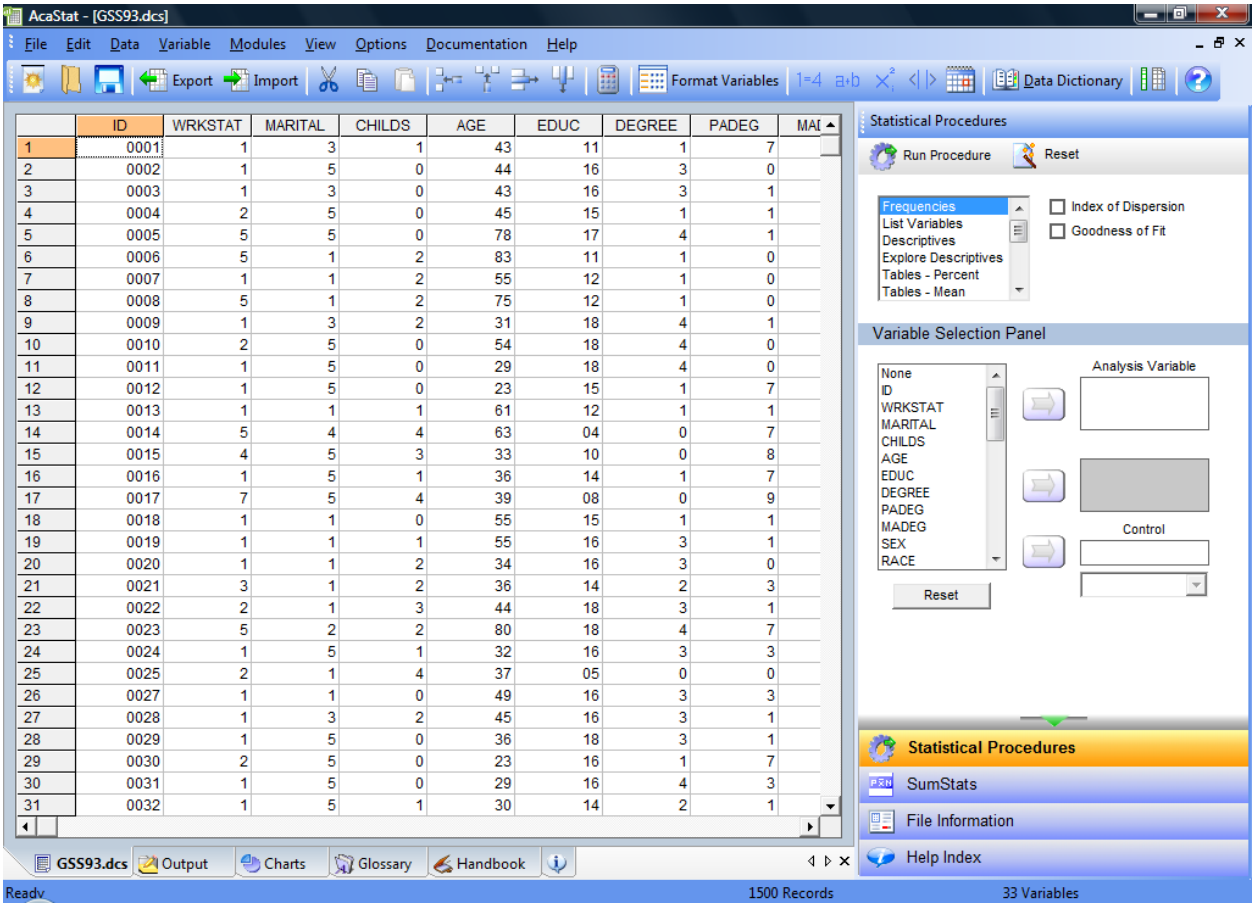






Data Grid

The Data Grid creates, imports, and exports delimited data files. The data spreadsheet displays data from a previously created data file or data you are currently entering into Data Grid. Opening or importing a new file will replace the current contents.



	ID	WRKSTAT	MARITAL	CHILDS	AGE	EDUC	DEGREE	PADEG	MA
1	0001	1	3	1	43	11	1	7	
2	0002	1	5	0	44	16	3	0	
3	0003	1	3	0	43	16	3	1	
4	0004	2	5	0	45	15	1	1	
5	0005	5	5	0	78	17	4	1	
6	0006	5	1	2	83	11	1	0	
7	0007	1	1	2	55	12	1	0	
8	0008	5	1	2	75	12	1	0	
9	0009	1	3	2	31	18	4	1	
10	0010	2	5	0	54	18	4	0	
11	0011	1	5	0	29	18	4	0	
12	0012	1	5	0	23	15	1	7	
13	0013	1	1	1	61	12	1	1	
14	0014	5	4	4	63	04	0	7	
15	0015	4	5	3	33	10	0	8	
16	0016	1	5	1	36	14	1	7	
17	0017	7	5	4	39	08	0	9	
18	0018	1	1	0	55	15	1	1	
19	0019	1	1	1	55	16	3	1	
20	0020	1	1	2	34	16	3	0	
21	0021	3	1	2	36	14	2	3	
22	0022	2	1	3	44	18	3	1	
23	0023	5	2	2	80	18	4	7	
24	0024	1	5	1	32	16	3	3	
25	0025	2	1	4	37	05	0	0	
26	0027	1	1	0	49	16	3	3	
27	0028	1	3	2	45	16	3	1	
28	0029	1	5	0	36	18	3	1	
29	0030	2	5	0	23	16	1	7	
30	0031	1	5	0	29	16	4	3	
31	0032	1	5	1	30	14	2	1	

Rows

Each row represents one record or case and is the unit of analysis for your procedures. If your data contain information on individuals, each row represents one person (observation). If your data contain information on cities, each row represents one city.

Columns

Each column represents data for all records regarding one variable or characteristic. If a column contains information on personal income in U.S. dollars, all data in that column should represent the income for each person (row) in the data file.

Cells

Each cell represents a value for one observation of a single variable. It can contain string data (words) or numerical data (numbers). Most data files use numerical data to represent words. As an example, for the variable sex, you could code Male as "0" and Female as "1". This makes data entry much easier since you only have to enter one number instead of typing a word. It also simplifies recoding variables and identifying controls.

Entering data

To enter data, point and click on a cell and begin typing. Pressing the Enter key will save the entry and move the data entry point to the next row. Pressing the right arrow key will save the entry and move the data entry point to the right. Pressing Escape (Esc) will cancel the data you are entering.

Editing data

To edit data, double-click on the cell and begin typing. Please note that pressing any of the arrow keys will save the entry and move the data entry point to the relevant cell. Pressing the Enter key will save the entry and move the data entry point to the next row. Pressing Escape (Esc) will cancel the data you are entering.

Right-Click Support

Click the right mouse button in the spreadsheet to select from several edit options to include adding and removing columns and rows and formatting variables.

Spreadsheet and Data File Limits

The Data Grid spreadsheet edit capacity is limited to 200,000 cells (more than sufficient for most analyses). This is calculated by multiplying the number of columns (variables) by the number of rows (observations). The edit capacity is also limited to 300 variables (columns).

Note: Although we strongly recommend you keep files under the edit capacity, you may import larger files but only the first 100 records will be displayed in the spreadsheet. If you import a file with more than 300 variables, the data grid will only display the first 300 variables, although all variables and observations will be available for analysis. You will not be able to edit larger files in the Data Grid, but can conduct recode operations and compute new variables. The file capacity will be limited by the amount of memory available on your computer. AcaStat has been tested on data files containing 50,000 observations and 200 variables. If you have large files, it is recommended that you reduce them by limiting the number of variables. A Split Data module is provided under the Data pull-down menu for creating duplicate files (same number of observations) with fewer variables.

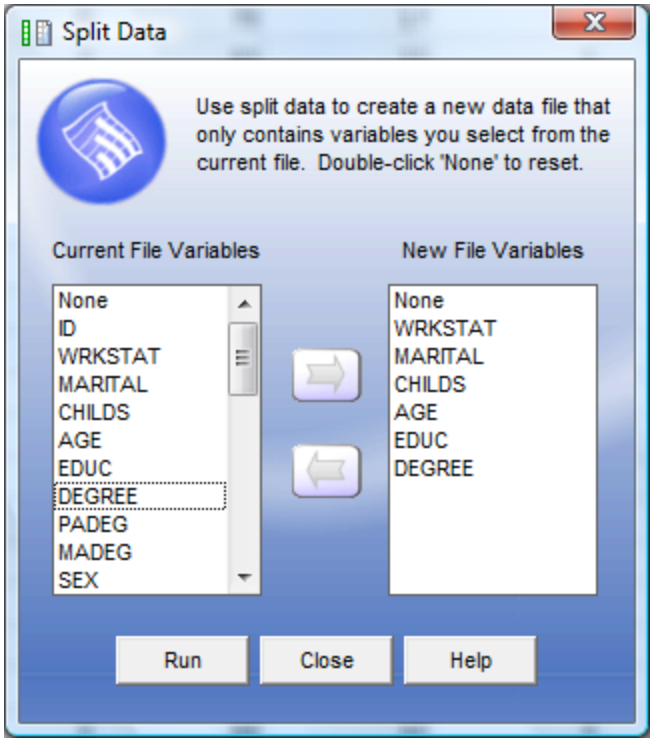
The table below gives some approximations of the maximum allowed when one considers the number of variables (Var) and observations (Obs) in one data file. Data sets exceeding these limits should be created in another program (we recommend spreadsheet software) and saved

as comma or tab delimited text files. Use the AcaStat import procedure to open the files for analysis.

Obs	Var	Obs	Var	Obs	Var	Obs	Var	Obs	Var
20000	10	3845	52	2127	94	1470	136	1123	178
16666	12	3703	54	2082	96	1448	138	1110	180
14285	14	3570	56	2040	98	1428	140	1098	182
12499	16	3447	58	1999	100	1407	142	1086	184
11110	18	3332	60	1960	102	1388	144	1074	186
9999	20	3225	62	1922	104	1369	146	1063	188
9090	22	3124	64	1886	106	1350	148	1052	190
8332	24	3029	66	1851	108	1332	150	1041	192
7691	26	2940	68	1817	110	1315	152	1030	194
7142	28	2856	70	1785	112	1298	154	1019	196
6666	30	2777	72	1753	114	1281	156	1009	198
6249	32	2702	74	1723	116	1265	158	999	200
5881	34	2631	76	1694	118	1249	160	989	202
5555	36	2563	78	1666	120	1234	162	979	204
5262	38	2499	80	1638	122	1219	164	970	206
4999	40	2438	82	1612	124	1204	166	961	208
4761	42	2380	84	1586	126	1189	168	951	210
4544	44	2325	86	1562	128	1175	170	942	212
4347	46	2272	88	1537	130	1162	172	934	214
4166	48	2221	90	1514	132	1148	174	925	216
3999	50	2173	92	1492	134	1135	176	916	218

Split Data Module

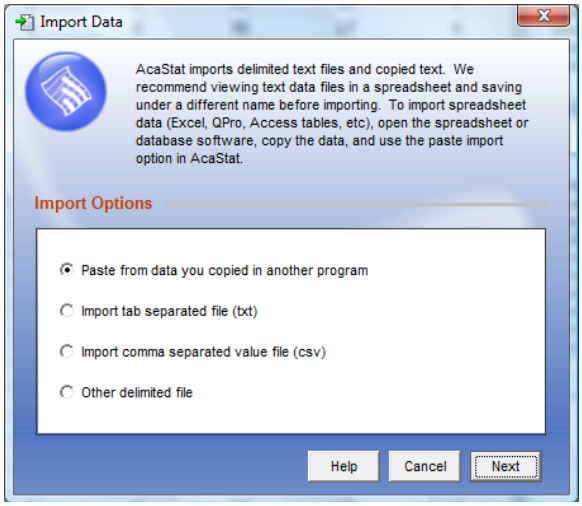
Use split data to create a new data file that only contains variables you select from the current file. This feature allows you to reduce the size of large data files to increase analysis speed and improve data editing capacity. The new file retains variable formatting.



To use the split data module, click the Data pull-down menu in the Data Grid tab and select the Split Data File option. A list of all the variables in the current data file will be provided. Select a variable from the current file list you wish to include in the new file and click the right arrow button. Continue adding variables until you are satisfied with the content of the new data file. To remove a variable from the new file list, select the variable and click the left arrow button. Double-click 'None' to reset the new file list.

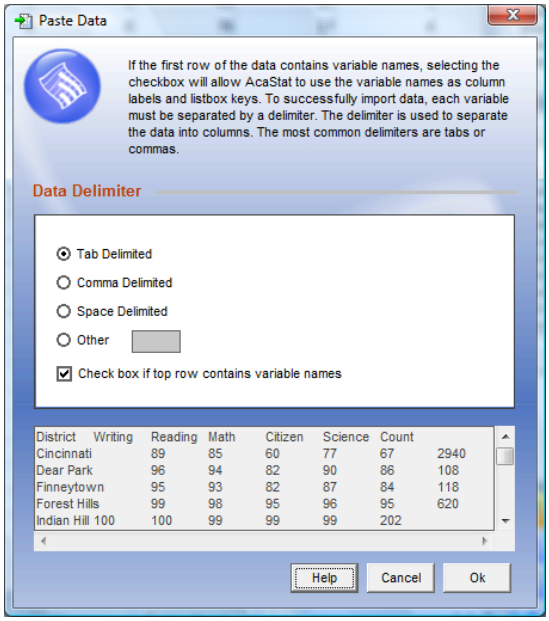
Importing Data

You can import data from text data files or by pasting data copied from a word processor, spreadsheet, or database table.


Import Data
AcaStat imports delimited text files and copied text. We recommend viewing text data files in a spreadsheet and saving under a different name before importing. To import spreadsheet data (Excel, QPro, Access tables, etc), open the spreadsheet or database software, copy the data, and use the paste import option in AcaStat.
Import Options
Paste from data you copied in another program
Import tab separated file (txt)
Import comma separated value file (csv)
Other delimited file
Help
Cancel
Next

Pasting Table Data

To paste data from a table, use your mouse to select and copy data in a word processor, spreadsheet, or database table. In Data Grid, select "File/Import/Paste Data". A Data Import module will display a portion of the data you copied and will provide options for indicating how the variables are delimited. Click OK to import the data. This replaces all data in the current Data Grid spreadsheet. You may also copy and paste individual cells, rows, and columns. Use the routine copy and paste functions for these procedures.


Paste Data
If the first row of the data contains variable names, selecting the checkbox will allow AcaStat to use the variable names as column labels and listbox keys. To successfully import data, each variable must be separated by a delimiter. The delimiter is used to separate the data into columns. The most common delimiters are tabs or commas.
Data Delimiter
Tab Delimited
Comma Delimited
Space Delimited
Other
Check box if top row contains variable names
District Writing Reading Math Citizen Science Count
Cincinnati 89 85 60 77 67 2940
Dear Park 96 94 82 90 86 108
Finneytown 95 93 82 87 84 118
Forest Hills 99 98 95 96 95 620
Indian Hill 100 99 99 99 202
Help
Cancel
Ok

Copying data from Microsoft Access

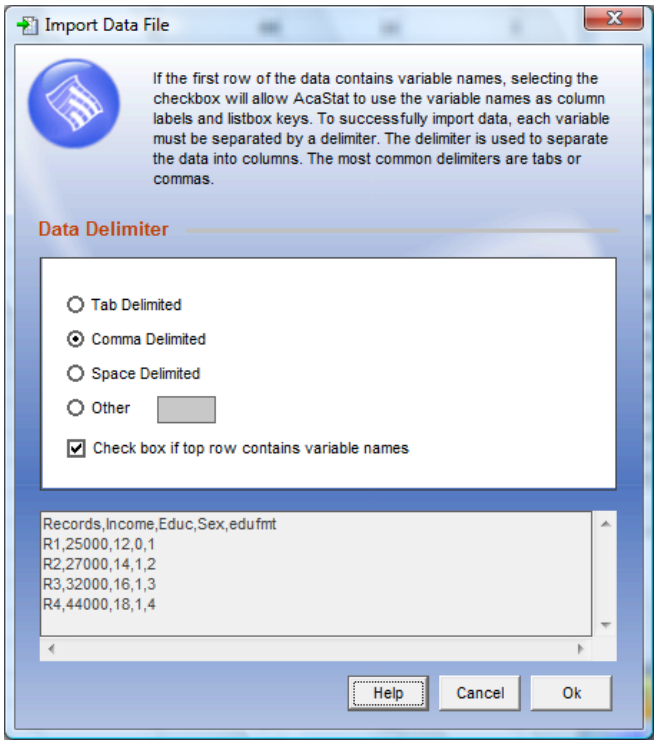
1. Click Table
2. Click Edit/Select All Records
3. Click Edit/Copy
4. Open Data Grid and use "File/Import/Paste Data"

Importing Text Data Files

To import a text file into Data Grid, click "File/Import Data" to select the data file. A Data Import module will display a portion of the data in the file and will provide options for indicating how the variables are delimited. You can create or export these text files from most software packages to include word processors, spreadsheets, databases, and statistical packages.

Data Grid primarily uses either comma separated value (csv) or tab separated value (txt) files, but you can use almost any delimiter. If the one you choose doesn't work, try another. If the first character in the file is not numerical, Data Grid will assume the first row in this file contains the variable labels and will analyze the data from row 2 on. You can change this by deselecting the checkbox.

If the data file is large, it may take a few moments for Data Grid to read all the data. Once read, the data will be displayed in the spreadsheet and variable names (column headings) will be visible in the list boxes, if they are contained in the first row of the data. Otherwise, Data Grid names the variables sequentially V1, V2, etc. You can edit data in the spreadsheet by double-clicking on a cell. After importing data, you should save the file as an AcaStat system file. This will increase loading speed and retains variable formatting.


Import Data File
If the first row of the data contains variable names, selecting the checkbox will allow AcaStat to use the variable names as column labels and listbox keys. To successfully import data, each variable must be separated by a delimiter. The delimiter is used to separate the data into columns. The most common delimiters are tabs or commas.
Data Delimiter
Tab Delimited
Comma Delimited
Space Delimited
Other
Check box if top row contains variable names
Records,Income,Educ,Sex,edufmt
R1,25000,12,0,1
R2,27000,14,1,2
R3,32000,16,1,3
R4,44000,18,1,4
Help
Cancel
Ok

Tips on using a spreadsheet application to create a csv or txt file

The first row in the spreadsheet should contain a column label to represent the data in that column (keep the labels 8 characters or less). These labels represent the variable names and help identify columns when imported into the Data Grid (see example of spreadsheet format below).

	A	B	C	D
1	AGE	SEX	INCOME	
2	23	M	46000	
3	45	F	55000	

Example of Import Error

When creating a comma separated value file, make sure all commas have been removed from the cell contents before creating the file. If they are not removed, Data Grid's import procedure will assume the comma is an extra variable. This will cause the cell contents in Data Grid to be out of alignment with the column. You can see below that data in rows two and three do not line up with the appropriate columns.

	AGE	SEX	INCOME	V4
1	65	M	33000	
2		23	M	
3	46000	45	F	

Exporting Data

You can export data to text data files or to a Microsoft Excel spreadsheet (Excel must be installed on your computer).

Sending Data to Excel

If you have Microsoft Excel software installed, select "File/Export Data/To Excel Spreadsheet". Excel will automatically open and Data Grid will populate the spreadsheet with the contents of the Data Grid. Don't forget to save the Excel file when done. This operation requires Excel97 or later versions. Large data file exports should be accomplished by saving to a text data file rather than a direct export to Excel.

Saving as a Text File

Click "File/Export Data" to save the data in the Data Grid datasheet as either a comma or tab separated (delimited) file. Exported files can be easily read by other software packages. Exporting data will loose any formatting you have created for variable labels, value labels, and missing values. If you open the exported file in a common spreadsheet, the first row will list the variable names and the remaining rows will contain the data (one row for each record). Each column represents the data for one variable.

SYLK Errors

Sometimes using a spreadsheet to open a text file exported from Data Grid (or from another software package) results in a SYLK error. If this occurs, try the following: a) open the file in a word processor or AcaStat's Output Log, b) select and copy the contents, and c) paste the contents into the spreadsheet. If you have Microsoft Excel software installed and the file is not large, use "File/Export Data/To Excel Spreadsheet".

Changing Variable Names

You may rename the variable names by clicking on any cell in the column you wish to rename and selecting the "Format" button on the toolbar (you may also use the "Variable Format" pull-down menu) or click the right mouse button to display a menu. This will open the format screen that will allow variable name changes. Keep variable names short by limiting them to 8 characters or less; e.g. "FamInc", not "Family Income."

Find Cell Contents

Use Find to search cell contents within one variable (column). This feature is most useful for finding specific observations or data. Cells with a match are bolded. Click find again to locate the next match. Click reset to remove the bolding.

Procedure

- ✓ Select the variable you wish to search by clicking on the variable name in the listbox.
- ✓ Enter the value or string you wish to search for.
- ✓ Click the "Find" button.

Find and Replace

Use find and replace to change cell contents within one variable (column). This feature is most useful for correcting data entry errors or changing string contents to values (e.g., to change “Male” and “Female” strings so that all “Male” strings are coded as the value 1 and all “Female” strings are coded as a value 2). An Undo button will appear after you conduct a replace operation. Clicking this button will return the cell contents to the immediate proceeding value. It is recommended that you conduct search and replace operations on a duplicate variable so you do not disturb the original variable. Use the Compute procedure to duplicate a variable.

Procedure

- ✓ Select the variable you wish to search by clicking on the variable name in the listbox.
- ✓ Enter the value or string you wish to search for.
- ✓ Enter the value you wish to use to replace the search parameter.
- ✓ Click the "Replace" button.

Adding/Deleting Rows and Columns

The data spreadsheet automatically begins new data files with 20 columns and 1000 rows.

Add/Delete Rows

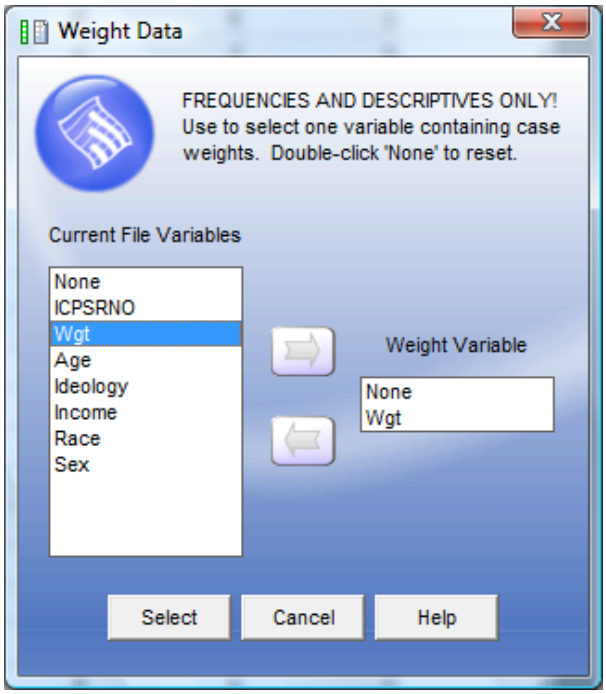
To add rows, click the Edit pull-down menu and select "Add Rows." Ten empty rows will be added to the bottom of the spreadsheet. To delete a row, select the rows you wish to delete, click the Edit pull-down menu, and select "Delete Rows." All data contained in the deleted row(s) are lost.

Add/Delete Columns

To add a column, click the Edit pull-down menu and select "Add Columns." One empty column will be added to the far right of the spreadsheet. To delete a column, select the columns you wish to delete, click the Edit pull-down menu, and select "Delete Columns." All data contained in the deleted column are lost.

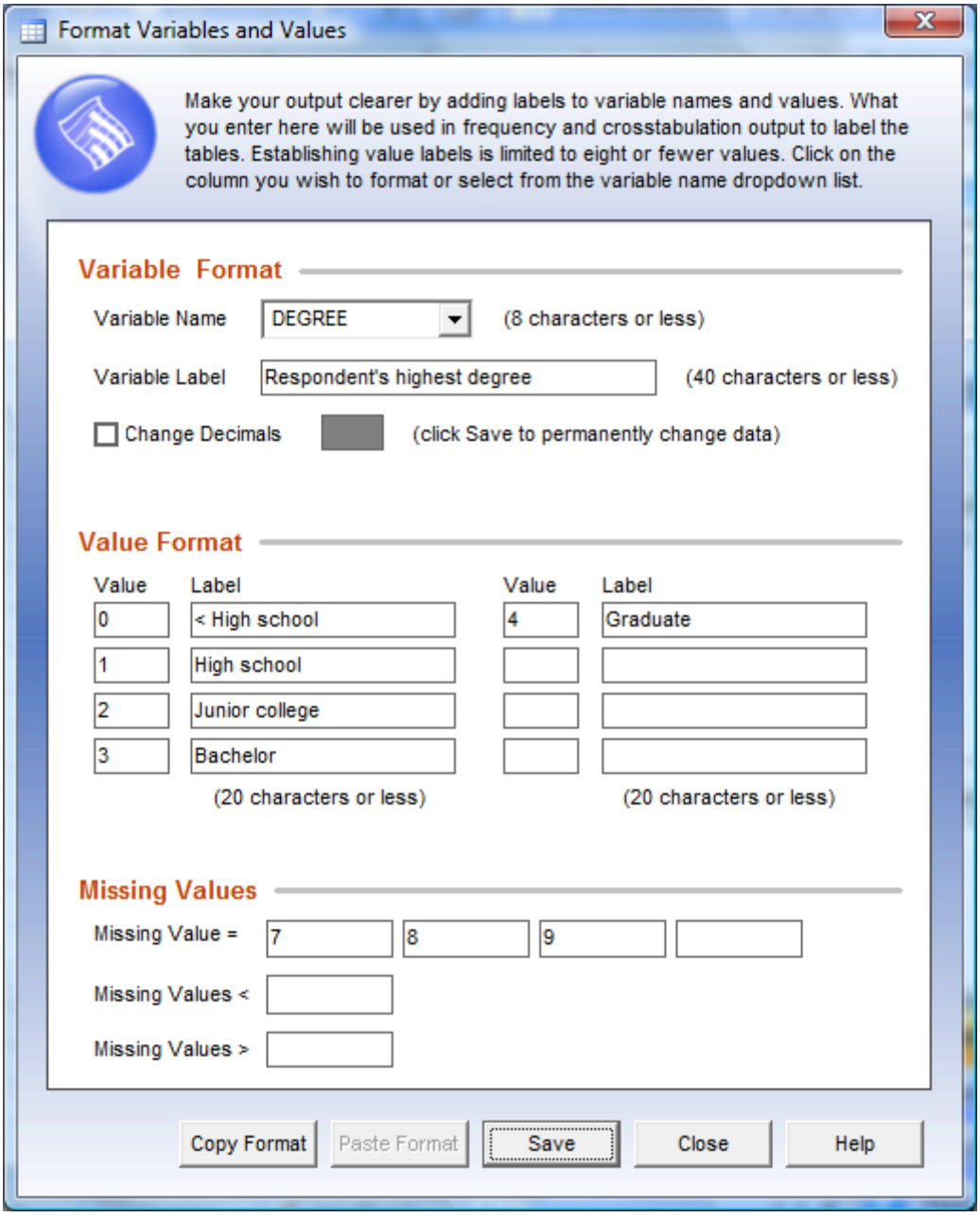
Using a Weight Variable

Click the Data pull-down menu and select 'Use Weight Variable' to identify a variable for weighting observations. Weights are often used to compensate for over or under sampling and also for producing population values from a random sample. The weights are used in the frequency and descriptive procedures. Weights will not be applied to any other procedures. Only one weight can be selected at a time.



Formatting Variables

To make output easier to interpret, create variable names and assign short explanatory labels to variable names and values. Formatting information will be saved with the AcaStat data file and can be changed at any time. Click the Copy Format button to copy and paste the formatting into another similarly coded variable (e.g., 1=yes and 2=no). If data are exported, the formatting information will not be exported.



Variable Name

Data Grid limits variable names to 8 characters. Since this is what is used to represent a column in the variable list boxes, it is important to create short but meaningful variable names. As an example, level of education could be named DEGREE or EDUC.

Variable Label

Use variable labels to create a more detailed description of the variable. The variable label is only used for creating output. Using the above example, DEGREE could be labeled "Respondent's highest degree."

Change Decimals

Specify the number of decimals displayed (and stored) for the selected variable. This option will permanently change the data file when the file is saved.

Value Labels

Use value labels to create a more detailed description of the values used in a variable. The value label is only used for creating output. If, for instance, education was coded 0 through 4 to represent five general categories, the values would be more meaningful if coded as the following: 0="< High School", 1="High school", 2="Junior college", 3="Bachelor", 4="Graduate". Without labels, the output would show the values 0 through 4.

Missing Values

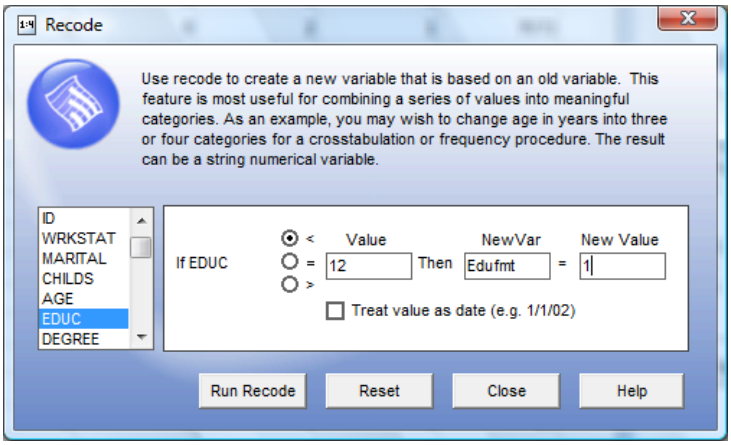
Identify up to four values or strings (words) to be excluded from the statistical analysis (blank cells are always considered missing). Values that are less than (<) or greater than (>) a set value can also be set as missing. Only numerical entries are useful for these two missing value conditions.

Recoding Values

Data Grid can use the values of one variable to create a new variable. This is most useful when there are several categories in a variable that can be combined into more meaningful subgroups. An example would be taking years of education and recoding the years into less than high school, high school, some college, and college graduate.

Date Values as Controls

Since statistical procedures require that all controls have numerical values, date values should be recoded into numerical values (e.g., a value of 1 might be used to represent all observations with a specific date, before a date, or after a date).


Recode
Use recode to create a new variable that is based on an old variable. This feature is most useful for combining a series of values into meaningful categories. As an example, you may wish to change age in years into three or four categories for a crosstabulation or frequency procedure. The result can be a string numerical variable.
ID
WRKSTAT
MARITAL
CHILDS
AGE
EDUC
DEGREE
If EDUC
Value
NewVar
New Value
12
Then
Edufmt
1
Treat value as date (e.g. 1/1/02)
Run Recode
Reset
Close
Help

Procedure

- ✓ Select the variable you wish to change by clicking on the variable name in the listbox.
- ✓ Select the operation you wish to perform by clicking on a radio button. For coding less than 12 years of education, select the "<" button.
- ✓ Enter the value you wish to use as the basis for the recode. For the above example, enter 12.
- ✓ Enter a variable name for the new variable you are creating (limited to 8 characters). Example "Edufmt" for formatted education.
- ✓ Enter the new variable's value that will be used to represent the old values. For our example, use 1 to represent anyone with less than 12 years of education.
- ✓ Click the "Run Recode" button.

A new column will be added to the Data Grid that has the assigned variable name and the recoded value. Repeat the process until all values needed for the new variable have been recoded.

Data Grid has been designed so the impact of each recode is visible. Using education as an example, any cell in the original education column with a value less than 12 should have the value of "1" on the same row in the new column labeled "EduFmt." Compare the education column to the EduFmt column to ensure the recode worked as planned.

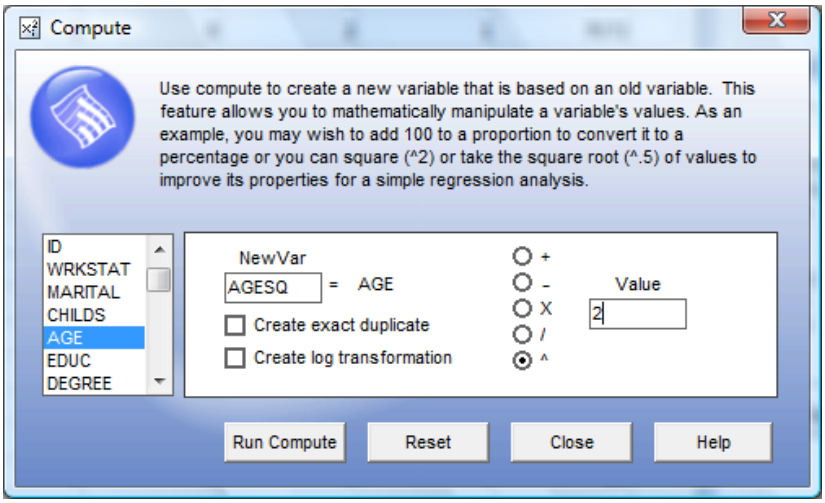
Computing New Variables

Data Grid can create a new variable by mathematically manipulating the values of the original variable. This is most useful when transforming a variable by squaring or taking the square root of the original values. It is also be useful for applying a constant to all observations in a variable to remove negative numbers or to develop an index. The Compute procedure can also be used to create an exact duplicate of a variable or create a new variable that represents the natural log of another variable's values.

Procedure

- ✓ Select the variable you wish to use for the compute from the variable name in the listbox.
- ✓ Enter a variable name for the new variable you are creating (limited to 8 characters). Example "AgeSq" for squaring age in years.
- ✓ Select the operation you wish to perform by clicking on a radio button.
- ✓ Enter the value you wish to use in the compute statement. Note that ^2=square and ^.5=square root.
- ✓ Click the "Run Compute" button.

A new column will be added to the Data Grid with the variable name and the computed values for all observations.

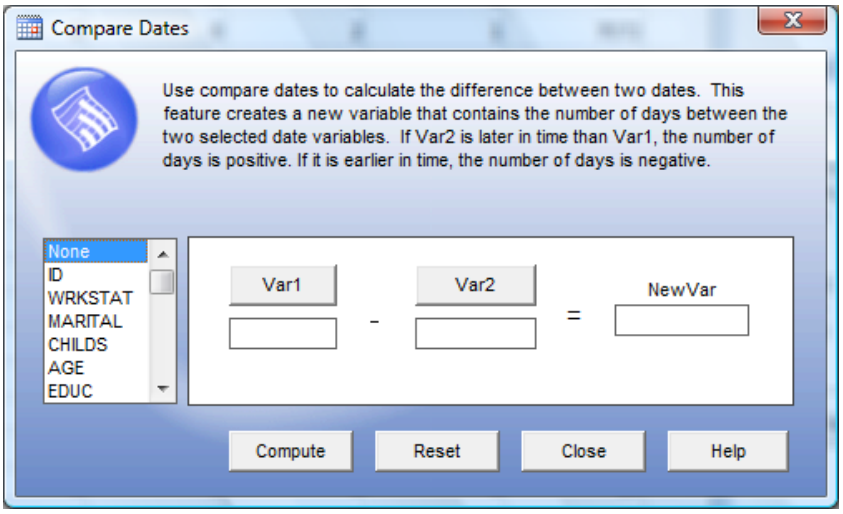


Comparing Dates

Use Compare Dates to calculate the difference between two dates. This feature creates a new variable that contains the number of days between the two selected date variables. If Var2 is later in time than Var1, the number of days is positive. If it is earlier in time, the number of days is negative.

Procedure

- ✓ Select the variable you wish to use for the first date from the variable name in the listbox and click the Var1 button.
- ✓ Select the variable you wish to use for the second date from the variable name in the listbox and click the Var2 button.
- ✓ Enter a variable name for the new variable you are creating (limited to 8 characters).
- ✓ Click the "Compute" button.

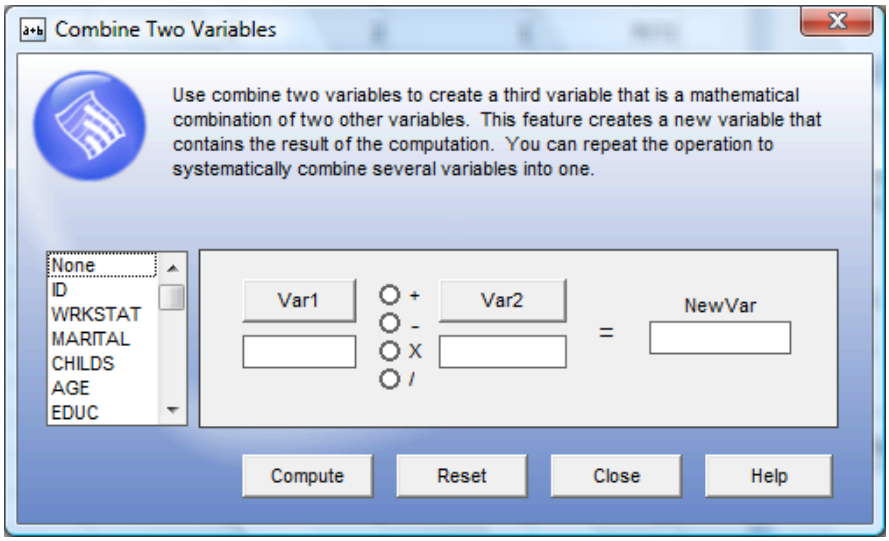

Compare Dates
Use compare dates to calculate the difference between two dates. This feature creates a new variable that contains the number of days between the two selected date variables. If Var2 is later in time than Var1, the number of days is positive. If it is earlier in time, the number of days is negative.
None
ID
WRKSTAT
MARITAL
CHILDS
AGE
EDUC
Var1
Var2
NewVar
-
=
Compute
Reset
Close
Help

Combining Two Variables

Use Combine Two Variables to create a third variable that is a mathematical combination of two other variables. This feature creates a new variable that contains the result of the computation. Repeat the operation to systematically combine several variables into one.

Procedure

- ✓ Select the first variable you wish to use from the variable name in the listbox and click the Var1 button.
- ✓ Select the second variable you wish to use from the variable name in the listbox and click the Var2 button.
- ✓ Enter a variable name for the new variable you are creating (limited to 8 characters).
- ✓ Select the operation you wish to perform by clicking on a radio button.
- ✓ Click the "Compute" button.

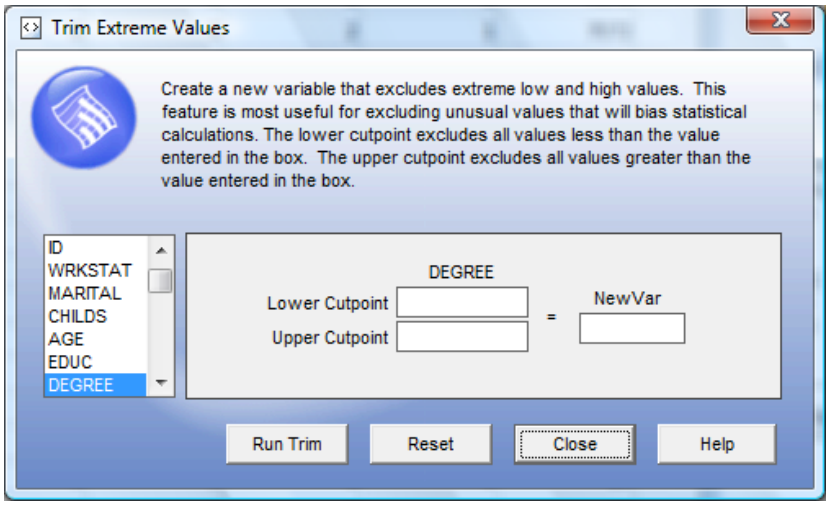


Trimming Extreme Values

Data Grid can create a new variable that excludes extreme observations. This is useful for excluding extreme values (outliers) from a variable that are biasing the statistical analysis. Trimming extreme values should be done cautiously, since meaningful information may be excluded with this operation. This caution is especially important for random samples where extreme values are an expected and necessary contribution to sample variation.

Procedure

- ✓ Select the variable you wish to trim from the variable name in the listbox.
- ✓ Enter the lower cutpoint. (excludes all values less than the entry)
- ✓ Enter the upper cutpoint. (excludes all values greater than the entry)
- ✓ Enter a variable name for the new variable you are creating (limited to 8 characters). Example "AgeTrim".
- ✓ Click the "Run Trim" button.


Trim Extreme Values
Create a new variable that excludes extreme low and high values. This feature is most useful for excluding unusual values that will bias statistical calculations. The lower cutpoint excludes all values less than the value entered in the box. The upper cutpoint excludes all values greater than the value entered in the box.
ID
WRKSTAT
MARITAL
CHILDS
AGE
EDUC
DEGREE
DEGREE
Lower Cutpoint
Upper Cutpoint
=
NewVar
Run Trim
Reset
Close
Help

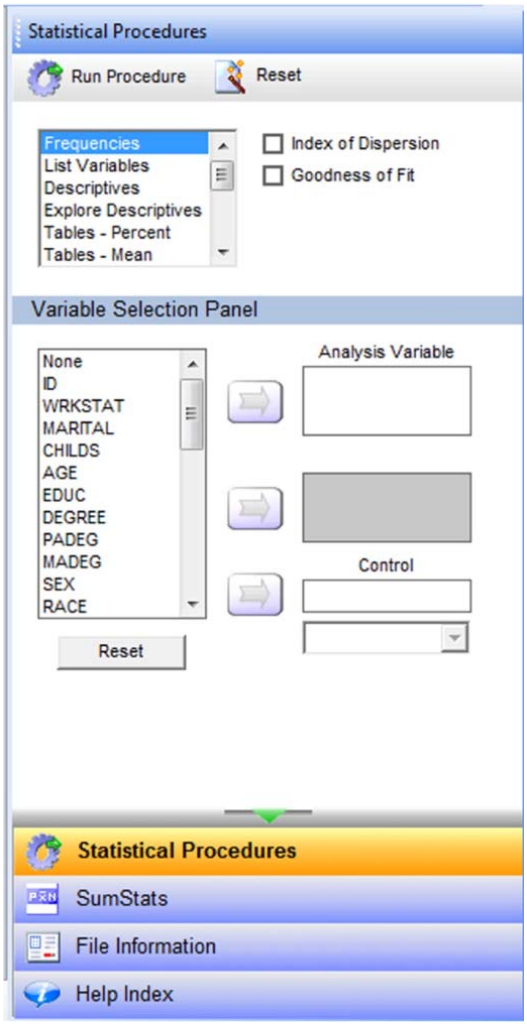
A new column will be added to the Data Grid that has the assigned variable name and blank cells where the extreme values would normally be located. All other cells will contain the data from the original variable.

Statistical Procedures

Use Statistical Procedures control panel to select variables and the type of analysis.

Variable Control Panel

Use the Variable Control Panel to select the variables you wish to use in the analysis. As an example, to produce frequencies tables for the variables sex and race, select the sex variable in the variable listbox and click the arrow button to place the variable in the "Analysis Variable" listbox. Use the same procedure to select and place the race variable in the listbox. Clicking the Run Procedure toolbar button will produce frequencies for both variables. Output from the analyses will be displayed automatically and recorded in the Output Log.



Dependent Variable

For any procedure, you must select an analysis or dependent variable. The topmost variable listbox is used for all univariate analyses (Frequencies and Descriptives).

Independent Variable

With the exception of the Frequencies, Descriptives, and List Variables procedures, an independent variable must also be selected (second listbox from top). In the case of t-tests, subgroup values must also be indicated for the independent variable so AcaStat compares the correct groups.

Control Variable

A control can be established for all procedures. This limits the analysis to a specific subgroup. As an example, if interested in the correlation between income (dependent variable) and education (independent variable) for females only, locate the Sex variable name from the variable listbox, click the arrow button to place the variable into the control box, and click the "=" button and indicate the value used for females. This would result in a correlation procedure that excludes all males. The control is normally a value (number) not a string (character) for ">", "<", and "<>" operators. If a string variable is used as a control, use the "=" operator or use the

Compute option (Variable Format pull-down menu) to create a duplicate of the variable and then use “Replace Cell Contents” (Edit pull-down menu) to replace strings with values.

The "All Missing" option limits the analysis to records that have missing data for the control variable. The "No Missing" option limits the analysis to all records that have valid entries for the control variable.

Note: Please remember to click the arrow button to remove the control variable before running another procedure.

Creating Data Subset with Controls

The control feature can be used to create a subset of the data based on the current file and the control settings. Select a control without any analysis variables and click the run procedure button. This will replace the current data with a subset that matches the control specification (such as only selecting cases where sex=1). This procedure replaces the current data in the Data Grid. Unsaved data will be lost (this operation does not affect saved files).

Statistical Procedures

You must select one of the statistical procedures to produce output.

Frequencies

This produces a listing of all the values in a variable, the number of times they occur in the variable, and the percent of cases represented in each value. Frequencies can be produced on numerical and string data.

Descriptives

Calculates measures of central tendency (mean, median) and variation (variance, standard deviation). Descriptives requires numerical data.

Explore

Explore descriptive statistics (sum, count, mean, sample standard deviation, standard error, 95% confidence interval) for each subgroup in the selected categorical variable(s). Explore requires numerical data.

Crosstabulation

This is a very useful procedure for representing the association between two categorical variables via a contingency table. Each cell in a contingency table represents a subgroup. For each cell, count, row percent, column percent, and total percent are reported. This procedure also automatically produces a chi-square statistic. Crosstabulations are easiest to interpret when the number of rows and columns in the contingency table are 5 or less.

Pearson Correlation

Calculates the Pearson correlation coefficient to represent the association between two interval/ratio level variables. For multiple comparisons, a correlation matrix can be display.

Regression

Calculates simple regression (one independent variable) or multiple regression (two or more independent variables). Ordinary least squares regression requires interval/ratio level dependent variables and logistic regression requires a dichotomous dependent variable. As an option, ordinary least squares regression will produce residuals for each observation and summary statistics for the observations included in the model. Logistic regression will produce predicted values and summary statistics.

List Variables

This procedure will list the values of observations for up to seven variables. List all the observations or set a maximum number of observations in the Max Obs box. As an example, if the data file had 1000 observations, a Max Obs value of 50 would list the first 50 observations in the data file.

Diagnostic Accuracy

Diagnostic accuracy compares a test result to a dichotomous reference standard. It is used to evaluate the sensitivity and specificity of a diagnostic test. The reference variable can only be two values (usually 0 for non-disease and 1 for disease). This represents the gold standard to compare to the test results. Each cell in the table holds counts. The test variable can be nominal, ordinal, or interval/ratio. The lowest value for the reference standard must always represent non-disease or the output will be incorrect. The cutpoint (T+ >) represents the point beyond which the test is positive (disease indicated). As an example, if possible test results range from 1 to 4, a cutpoint of 1 means a positive test is indicated by a value of 2, 3, or 4. You should carefully evaluate the sensitivity and specificity for one cutpoint to ensure you have configured the data properly before interpreting the results.

Definitions of abbreviations used in the output

Sn Sensitivity:

The proportion of diseased cases with a positive test.

Sp Specificity:

The proportion of non-diseased cases with a negative test.

PPV Positive Predictive Value:

The proportion of cases with a positive test that have the disease.

NPV Negative Predictive Value:

The proportion of cases with a negative test that do not have the disease.

LR+ Positive Likelihood Ratio:

A positive ratio greater than 1 indicates the diagnostic test increases the odds of identifying disease Sn/(1-Sp).

LR- Negative Likelihood Ratio:

A negative ratio less than 1 indicates the diagnostic test increases the odds of correctly identifying disease absence (1-Sn)/Sp.

K Kappa:

The level of agreement beyond chance for a 2x2 table at the cutpoint.

T-tests of means

Used to test for the difference between two sample means. For t-tests, the dependent variable must be an interval/ratio level variable to produce the means for analysis. The independent variable must be a nominal or ordinal level variable (often it will be a dichotomous variable such as sex: 1=male, 2=female). Values representing the comparison groups for the independent variable must be entered. These values are used to separate the dependent variable into two groups. Be careful to use the correct values or there may be an error reported that indicates there are not enough cases to compute the t-test. If there are more than two values in the independent variable, you may select the two subgroups you are interested in testing. The other subgroups will be excluded from the analysis.

Analysis of Variance (ANOVA)

Used to test for the difference between two or more sample means. Like t-tests, the dependent variable must be an interval/ratio level variable to produce the means for analysis. The independent variable must be a nominal level variable (e.g.: race: 1=white, 2=black, 3=Asian, 4=Hispanic. or ordinal (political ideology: 1=conservative, 2= moderate, 3=liberal). ANOVA optional output includes summary statistics for each subgroup and the Bonferroni post hoc test for multiple comparisons. Charts are not provided for the ANOVA procedure.

Summary Report

Summary report creates a table of summary statistics that can be printed or exported. This is a very useful option for displaying survey data and market research results in a format more understandable for those not use to interpreting contingency tables or standard descriptive statistics output. The categorical option will produce counts and column percents for each category contained in the columns by each category available in the rows. For continuous data, the report will produce summary statistics for each row variable when controlling for the column category. More than one column variable can be used but it is important to select column variables that are categorical, otherwise the table may be too large to print or interpret.

Summary reports are sensitive to missing data, so carefully review the data before and after analysis. It is especially important that the first dependent variable (rows variable) contain data for all the values of the independent variable (columns categories). If this is an issue, use the following procedure to create a placeholder variable:

1) Add a blank column to the spreadsheet

2) Right click a cell in the column and select “Replace Cell Contents” (also under the Edit menu)

3) Click the list name of the blank column you created and enter the value 1 in the field labeled “Replace with this value or string”. Click the Replace button and then close

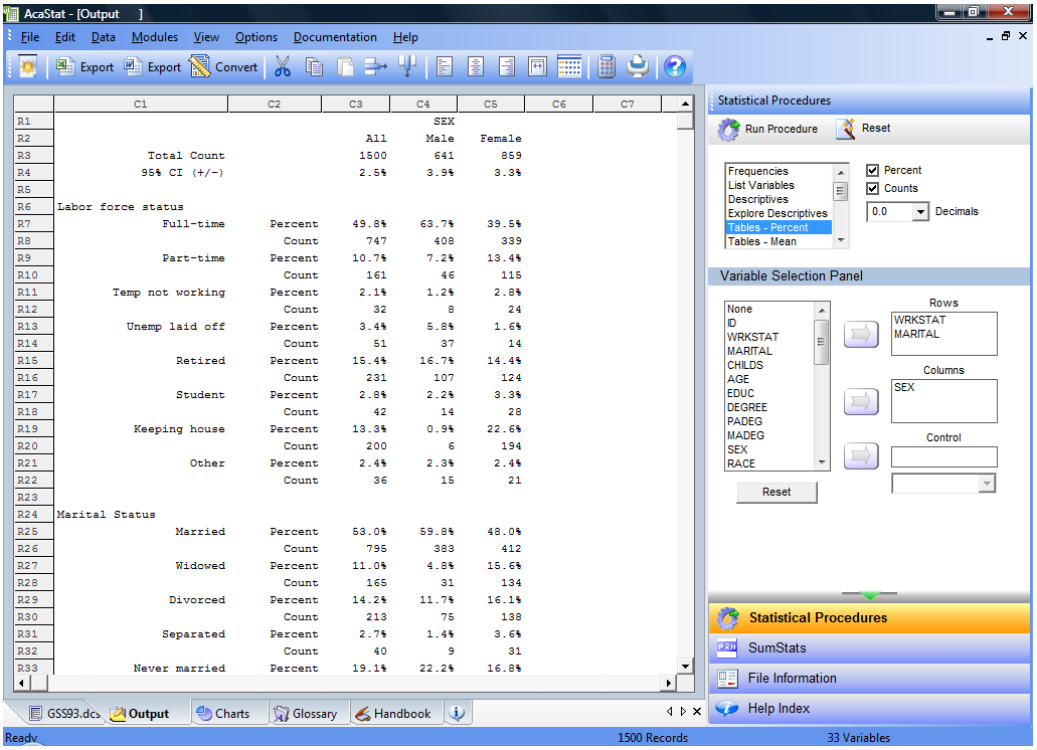
4) When setting up the variables for the summary reports, select the placeholder variable you created as the first row variable. Then add the row variables you wish to analyze.

5) After running the procedure, delete the placeholder variable from the report.

When a summary report is created, the Analysis Output tab will display the results in a spreadsheet. Edit the data in the spreadsheet, format column widths, delete rows and columns, and edit headers and footers for printed output. The summary report can also be exported to Word or Excel or copied and pasted into other software. The spreadsheet allows right-click mouse support.

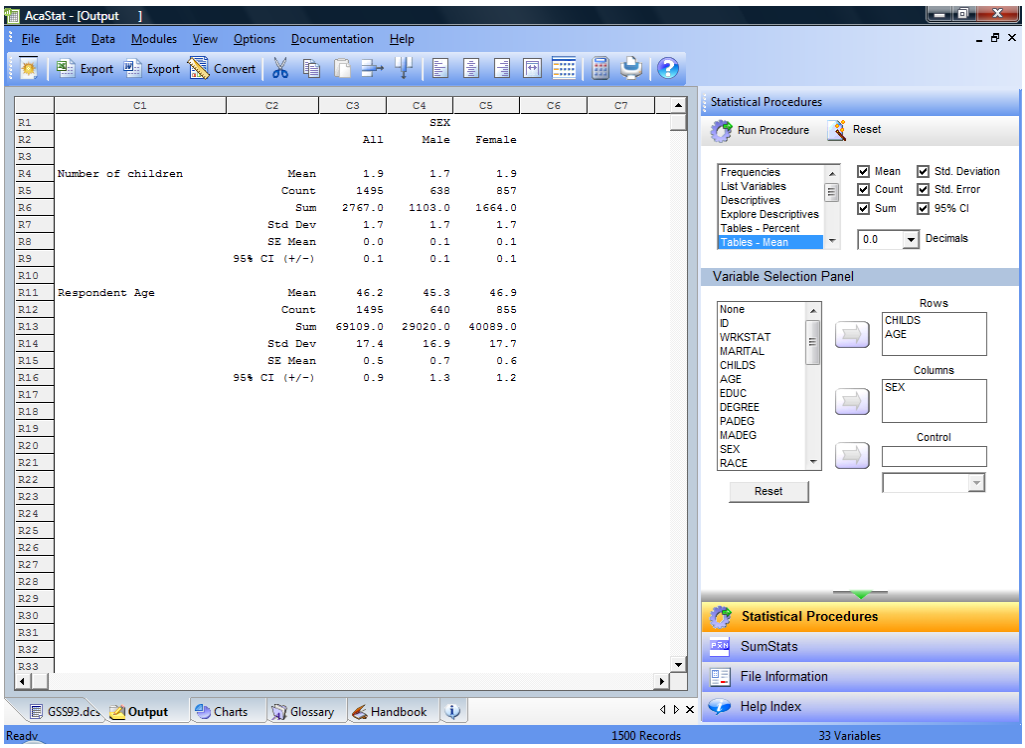
Note: The summary report procedure is the only procedure that does not append the results to the Output Log unless it is converted to tab output (Edit menu). This conversion, where tabs are inserted in place of the spreadsheet columns, is appended to the output log.

Summary Report for Categorical Data



	C1	C2	C3	C4	C5	C6	C7
R1				SEX			
R2			All	Male	Female		
R3	Total Count		1500	641	859		
R4	95% CI (+/-)		2.5%	3.9%	3.3%		
R5							
R6	Labor force status						
R7	Full-time	Percent	49.8%	63.7%	39.5%		
R8		Count	747	408	339		
R9	Part-time	Percent	10.7%	7.2%	13.4%		
R10		Count	161	46	115		
R11	Temp not working	Percent	2.1%	1.2%	2.8%		
R12		Count	32	8	24		
R13	Unemp laid off	Percent	3.4%	5.8%	1.6%		
R14		Count	51	37	14		
R15	Retired	Percent	15.4%	16.7%	14.4%		
R16		Count	231	107	124		
R17	Student	Percent	2.8%	2.2%	3.3%		
R18		Count	42	14	28		
R19	Keeping house	Percent	13.3%	0.9%	22.6%		
R20		Count	200	6	194		
R21	Other	Percent	2.4%	2.3%	2.4%		
R22		Count	36	15	21		
R23							
R24	Marital Status						
R25	Married	Percent	53.0%	59.8%	48.0%		
R26		Count	795	383	412		
R27	Widowed	Percent	11.0%	4.8%	15.6%		
R28		Count	165	31	134		
R29	Divorced	Percent	14.2%	11.7%	16.1%		
R30		Count	213	75	138		
R31	Separated	Percent	2.7%	1.4%	3.6%		
R32		Count	40	9	31		
R33	Never married	Percent	19.1%	22.2%	16.8%		

Summary Report for Continuous Data



	C1	C2	C3	C4	C5	C6	C7
R1				SEX			
R2			All	Male	Female		
R3							
R4	Number of children	Mean	1.9	1.7	1.9		
R5		Count	1495	638	857		
R6		Sum	2767.0	1103.0	1664.0		
R7		Std Dev	1.7	1.7	1.7		
R8		SE Mean	0.0	0.1	0.1		
R9		95% CI (+/-)	0.1	0.1	0.1		
R10							
R11	Respondent Age	Mean	46.2	45.3	46.9		
R12		Count	1495	640	855		
R13		Sum	69109.0	29020.0	40089.0		
R14		Std Dev	17.4	16.9	17.7		
R15		SE Mean	0.5	0.7	0.6		
R16		95% CI (+/-)	0.9	1.3	1.2		
R17							
R18							
R19							
R20							
R21							
R22							
R23							
R24							
R25							
R26							
R27							
R28							
R29							
R30							
R31							
R32							
R33							

Simulation

Two simulation procedures can be used for manipulating data in the spreadsheet. They were designed to help students develop a better understanding of random sampling. The Random Sample option creates a random sample from a larger data file. If you assume the larger data file represents the entire population, a smaller random helps students discover how summary sample measures such as means and proportions will vary from population parameters.

The Repeated Random Sampling procedure conducts multiple random samples to build a data set that reflects summary measures for each random sample. This procedure helps students test the Central Limit Theorem:

Central Limit Theorem

> As sample size increases, the sampling distribution of means approximates a normal distribution and is close to normal at a sample size of 30.

Random Sample

This procedure is used to create a data file that is a randomly selected subset of the current spreadsheet. Enter the number of observations desired. Please note that the number of observations must be less than total number of records in the current data. Click CREATE SAMPLE to replace the current data with a random sample.

Repeated Random Sampling

This procedure creates repeated random samples from the current spreadsheet. For each sample (iteration), the mean and standard deviation are computed and placed into the spreadsheet. This results in a new data file where each row represents the summary statistics for one random sample.

Enter the number of observations (sample size) and the number of random samples (iterations) desired. Please note that the sample size must be less than the total number of records in the current data. Click RUN ITERATIONS to replace the current data with a sampling distribution.

Appraisal Ratios

Use the Module pull-down menu to set which statistical procedures are available. The Statistics Module is the default and provides all the statistical analyses available in AcaStat except Appraisal Ratios. The Appraisal Module provides an additional procedure that is used to determine tax assessment values of real estate. Selection of the Appraisal Module also changes the contents of the Glossary and the output of certain procedures.

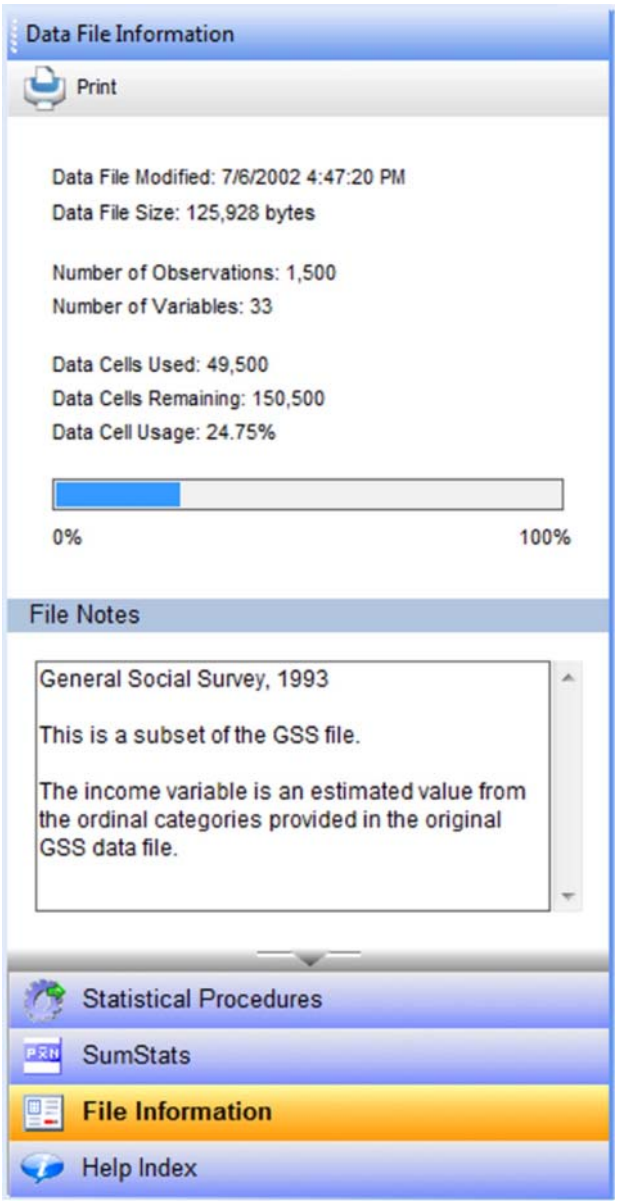
When the Appraisal Module is selected, the Appraisal Ratios statistical procedure will be visible in the procedure list. Use of this procedure requires data on the market (selling) price and the assessment value of homes to compute ratios and other statistics. A practice data file “Tax Assessments” is included with AcaStat.

Some statistical procedures are hidden when the Appraisal Module is active. To access these procedures, re-activate the Statistics Module through the Module pull-down menu.

Project Notes

The Project Notes and Data File Information panel provides information to help monitor the size of the data file and the contents of the data grid (spreadsheet). There is also a text box for keeping brief project notes and comments.

Use the Notes textbox to keep project and progress notes. The notes are saved with the data file. To limit the impact on the size of the data file, the notes field is limited to 3000 characters (about one 8.5 x 11 page of text). Additions or changes to the notes will be lost if the data file is not saved before exiting. Since saving Project Notes also saves the contents of the current data file, make sure the current configuration of the spreadsheet is acceptable. When in doubt, simply create a different file name when saving. Click the print button to print the project notes.


Data File Information
Print
Data File Modified: 7/6/2002 4:47:20 PM
Data File Size: 125,928 bytes
Number of Observations: 1,500
Number of Variables: 33
Data Cells Used: 49,500
Data Cells Remaining: 150,500
Data Cell Usage: 24.75%
0%
100%
File Notes
General Social Survey, 1993
This is a subset of the GSS file.
The income variable is an estimated value from the ordinal categories provided in the original GSS data file.
Statistical Procedures
SumStats
File Information
Help Index

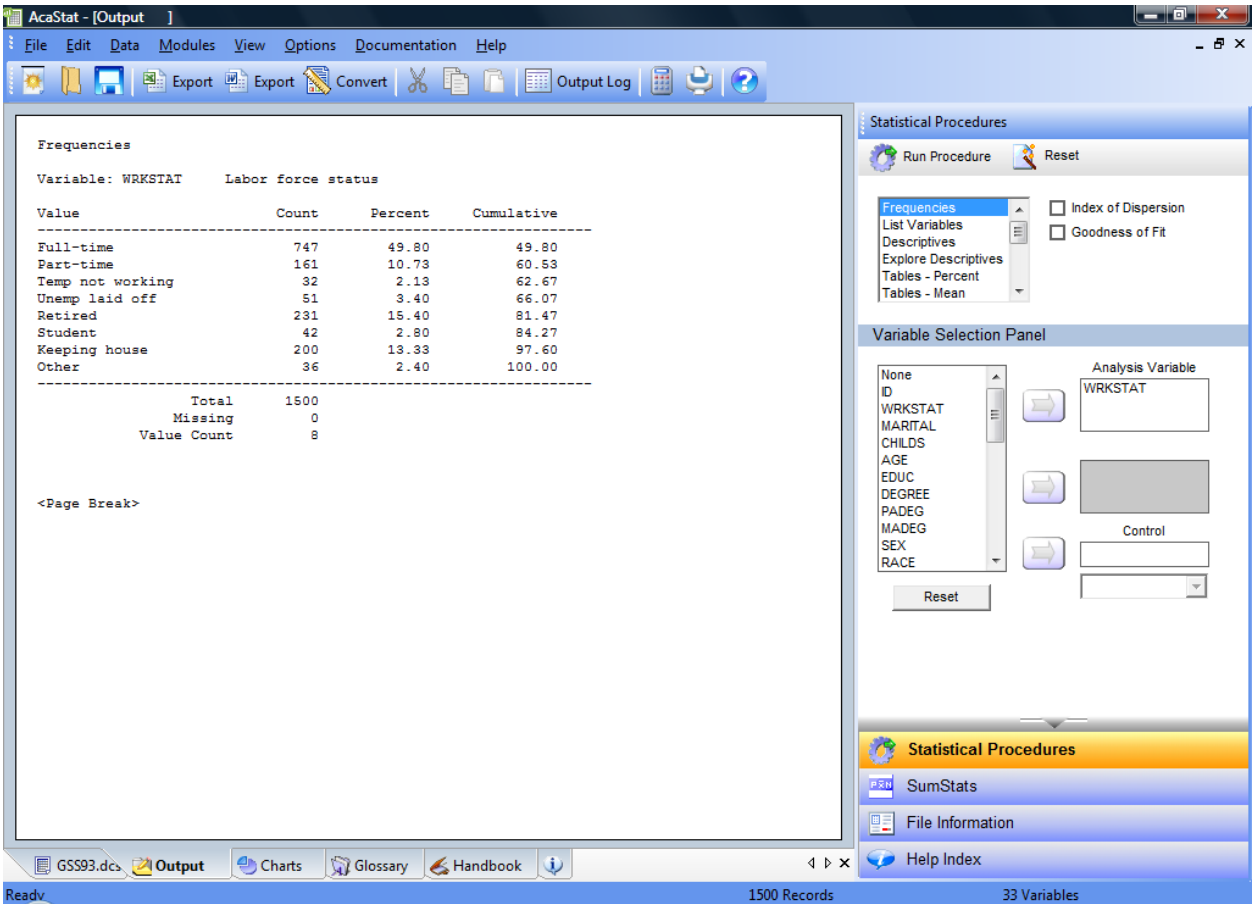
Data File Information

The Data File Information control panel contains information about the spreadsheet and current data file. If using a saved AcaStat data file, information is provided on when the data file was last modified and file size. Information is also provided on the number of variables and observations as well as the remaining capacity of the spreadsheet.

Output

The Output tab displays the results of the analysis in a text window for editing, printing, and/or saving as a text file. Each time a statistical procedure is run, the results are automatically sent to the output log. Each subsequent transfer of results will be appended to previously appended results creating a complete record of the analyses. The Output tab can also be used to view and create csv data files. Closing the output log does not erase the contents. Click the "Clear" button to erase all output. If Microsoft Word installed on the host computer, click the Word icon on the toolbar to export the output to Word. This procedure automatically formats the output and places each analysis on a separate page.

There are two output view options. Most Recent Output displays only the results of the most recent statistical procedure. Output log displays the output from multiple statistical procedures. It is important to remember that the output log contains all the analyses created, so the most recent analysis is at the end of the log. The toolbar displays the output option available if clicked, so when viewing Most Recent Output the toolbar button will display “Output Log” to indicate clicking the button will display the entire output. In turn, when viewing Output Log, the toolbar button will display the caption “Most Recent Output.”



Page Breaks

A page break will be inserted after each procedure and is marked in the document as "<page break>". This is used to indicate where a new page should be printed when printing the results. Deleting the "<page break>" notation will stop a new page from being printed. Adding this notation will cause AcaStat to print a new page at the point where you inserted the notation.

Printing

By default, AcaStat prints a date and time header and a Notes section on each page printed. Use the Options pull-down menu to deactivate the default settings.

Exporting

The contents of the Output tab can be exported to Microsoft Word (Word97 and newer) by clicking on the Word icon in the toolbar. The contents of the Output tab can also be exported to Microsoft Excel (Excel97 and newer) by clicking on the Excel icon in the toolbar.

Copying and Pasting Output

The Output tab uses a Courier New font. When copying and pasting the results into another document (which we recommend for editing and including with a report), make sure the font of the output is set to Courier New or the tables may not align properly. Some word processing software will format the pasted text to match that software's default font setting. This will cause the pasted text to become misaligned.

It is important to remember that the output log contains all the analyses created, so scroll to find the most recent analysis.

If the Summary Report procedure is used, the Output Tab will display a spreadsheet table. Edit cell contents, format the width of columns, format the header and footer for printing, copy and paste into other software, or export to Microsoft Word or Excel. Right-click mouse support is provided. The summary report procedure is the only procedure that does not append the results to the output log unless it is converted to tab delimited output (Edit menu). This conversion is appended to the output log and is best used for copying and pasting output into a spreadsheet.

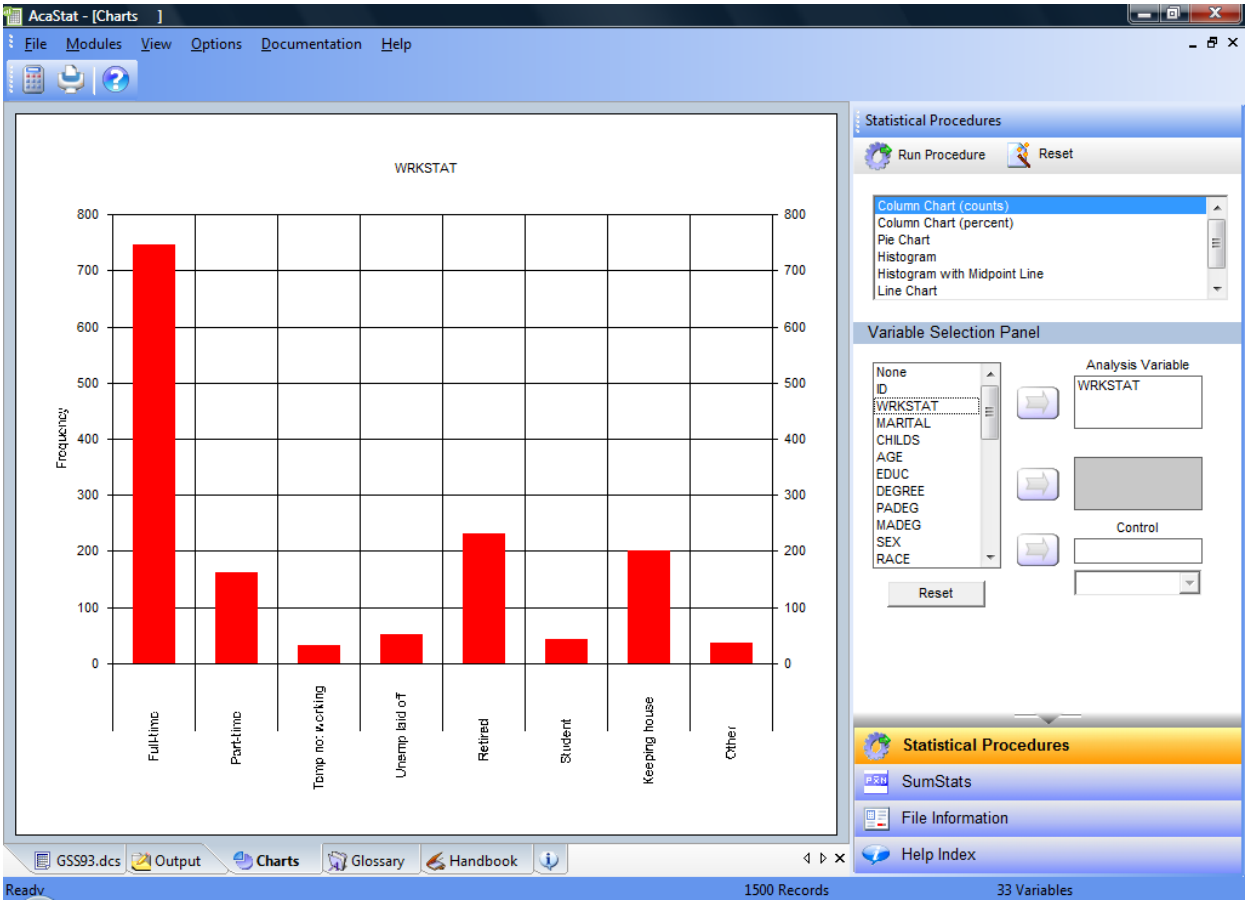
Converting Output to Tab Delimited

This option located under the Edit menu will convert all output tables into a tab delimited format. The conversion improves copying and pasting output into spreadsheets such as Excel or Quattro Pro.

Charts

Use the Charts tab to produce histograms, scattergrams, column, and pie charts. Charts are produced for single comparisons (one dependent variable and, when relevant, one independent variable).

If a chart is visible, click the chart. This will open the chart in a separate form to enlarge, copy, and print. Resize the chart form to enlarge or shrink an image (this may impact the axis label formatting). If pasting the graph into other software, copy the chart and then use the "Edit/Paste Special" pull-down menu in the software you wish to paste the graph. Select the "bitmap" option.

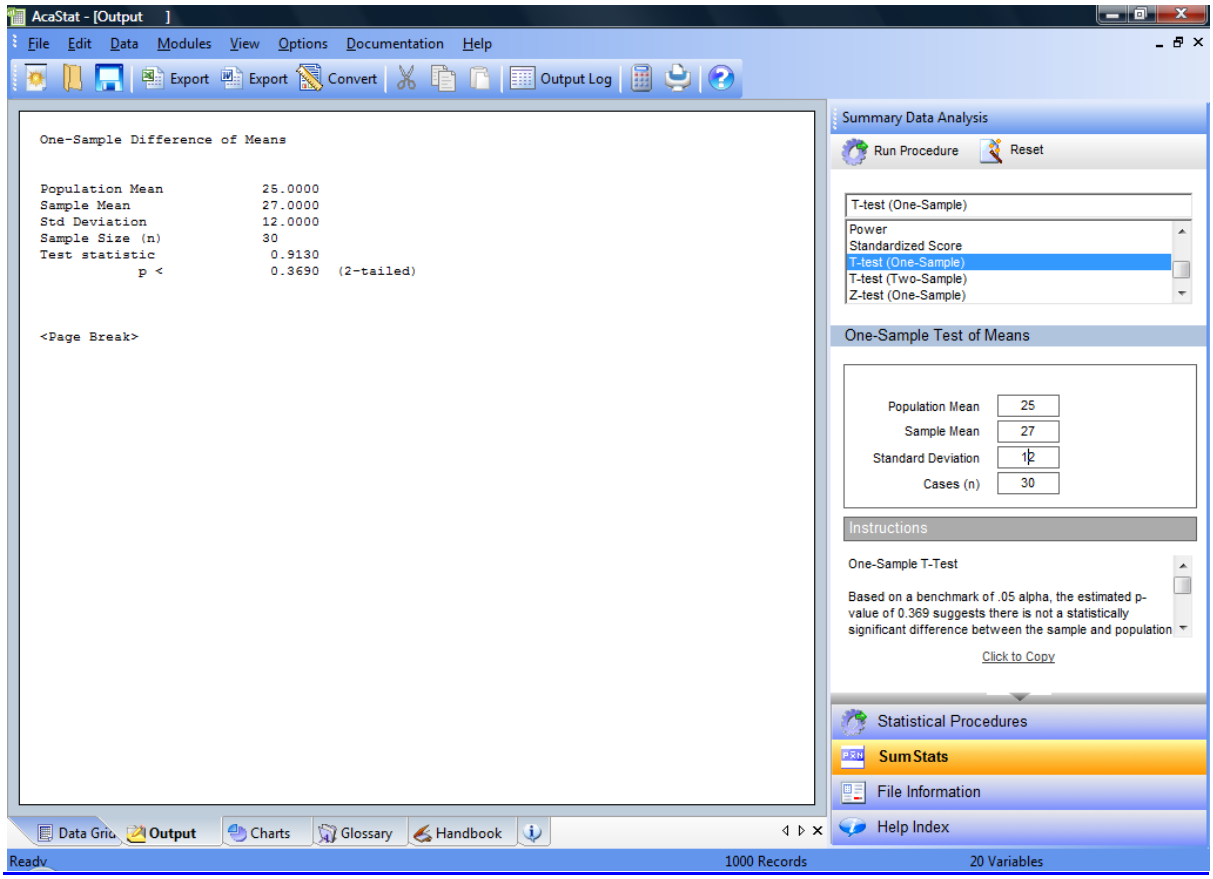


SumStats is designed for analyzing summary statistics. Summary statistics are developed from raw data by programs such as AcaStat, SAS, and SPSS. Without access to the raw data, it is difficult to conduct statistical comparisons of means, counts, and percentages. SumStats solves this problem by allowing the use of summary data to calculate basic significance tests and other common descriptive calculations.

Most statistics courses use summary statistics to teach the fundamentals. SumStats is very useful for checking the results of hand calculations.

SumStats Tips

- Module specific instructions are presented as you click on a new module.
- Instructions are replaced with summarized results from a calculation.
- Use the ENTER key to move between boxes.
- Pressing the ENTER key in the last box requiring data automatically calculates results.
- Analyses are automatically appended to the Output Log.
- Use copy and print buttons to print or paste the analysis results for the current procedure.
- Entering a label is not required but is useful for interpreting output.



t-test Module

One Sample Mean

To use the t-test module for one sample mean, enter the data in the boxes as indicated and click the calculate button. The population mean is the hypothetical mean that is used to compare to the sample mean. If left blank, SumStats will assume it is 0.

This module may also be used for paired t-tests. For a paired t-test, collect the difference for each case between Var1 and Var2 and calculate the mean and standard deviation for this difference. Enter this information into the relevant One Sample Mean boxes and enter 0 for the population mean. This will test the null hypothesis that the paired difference between the means of Var1 and Var2 is not significantly different from zero.

Two Sample Means

To use the t-test module for two sample means, enter the data in the boxes as indicated and click the calculate button. P-values will be provided for samples with equal and unequal variance.

ANOVA Module

Enter the data in the boxes as indicated and click the Run Procedures button. Each column represents one sample distribution. You must start with the left column. At least two columns (the two left columns) must have data for SumStats to compute statistics.

Z-test Module

One Sample Proportion

To use the Z-test module for one sample proportion, enter the data in the boxes as indicated and click the calculate button. The population proportion is the hypothetical proportion that is used to compare to the sample proportion. If left blank, SumStats will assume it is 0.

Two Sample Proportions

To use the Z-test module for two sample proportions, enter the data in the boxes as indicated and click the calculate button. It is important to remember that, when comparing percentages, first convert them to proportions (e.g., 75% = .75).

Chi-square Module

To use the Chi-square module, enter frequencies (counts) in the boxes and click the calculate button. The columns represent one variable with up to five subgroups (e.g., Protestant, Catholic, and Muslim are subcategories for the variable "Religion"). The rows represent a second variable with up to five subgroups. Start with the left column. At least two columns (the two left columns) and two rows (the top two rows) must have data for SumStats to compute statistics.

Margin of Error for Means Module

Enter the sample size (n), mean, and standard deviation (std dev) and click the calculate button to estimate the margin of error. SumStats will produce 90%, 95%, and 99% confidence intervals (in order from top to bottom).

Margin of Error for Proportions Module

Enter the sample size (n) and the proportion and click the calculate button to estimate the margin of error. SumStats will produce 90%, 95%, and 99% confidence intervals (in order from top to bottom).

Standardized Score Module

To use the standardized score module, enter one score from a distribution of scores in the first box. Enter the mean of the distribution in the second box. Enter the standard deviation of the distribution in the third box. Click the calculate button and the z-score will be calculated and its relative location in a normal distribution will be noted in the graph.

Diagnostic Accuracy Module

This module is used to evaluate the accuracy of a diagnostic test in identifying disease presence or absence. The 2x2 grid uses counts from a crosstabulation between known disease status and the diagnostic test results for that disease for each patient. To use this module, enter the counts in the relevant boxes and click Run Procedure.

If the counts are not already provided, counts can be produced by:

a) creating a data file in Data Grid where the rows represent patients and there is one column for known disease status (0=No, 1=Yes) and one column for test results (0=Neg, 1=Pos);
b) running the crosstabulation procedure where the test variable (column) is the dependent variable and the disease status is the independent variable; and,
c) entering the counts from the Crosstabulation into this module.

There is also a Diagnostic procedure in Data Grid for creating diagnostic accuracy statistics using a raw data file.

Power Module

Four criteria are used to estimate the appropriate sample size for a study. Sometimes called a power analysis, the primary goal is to collect enough cases to ensure you don't make a Type II error (Beta). Beta is the probability of not rejecting a null hypothesis when it should have been rejected. Power is 1- beta and is defined as the probability of correctly finding statistical significance. A common value for power is .80 or an 80% chance that a random sample will find a statistically significant difference when there truly is a difference in the population. SumStats is focused specifically on 2-tailed tests between two independent random samples (means and proportions). There are many other approaches and equations for estimating sample size.

To use the power module, select the statistical significance (alpha) and statistical power (1-Beta). If you are comparing two proportions, use the z-test sample boxes. Enter the expected proportion for each sample and click the calculate button. If you are comparing two means, use the t-test sample boxes. Enter the expected difference between the two sample means and the estimated pooled variation (standard deviation) and click the calculate button. SumStats will present the total sample size needed given your assumptions. Note that the sample size is the total of both samples.

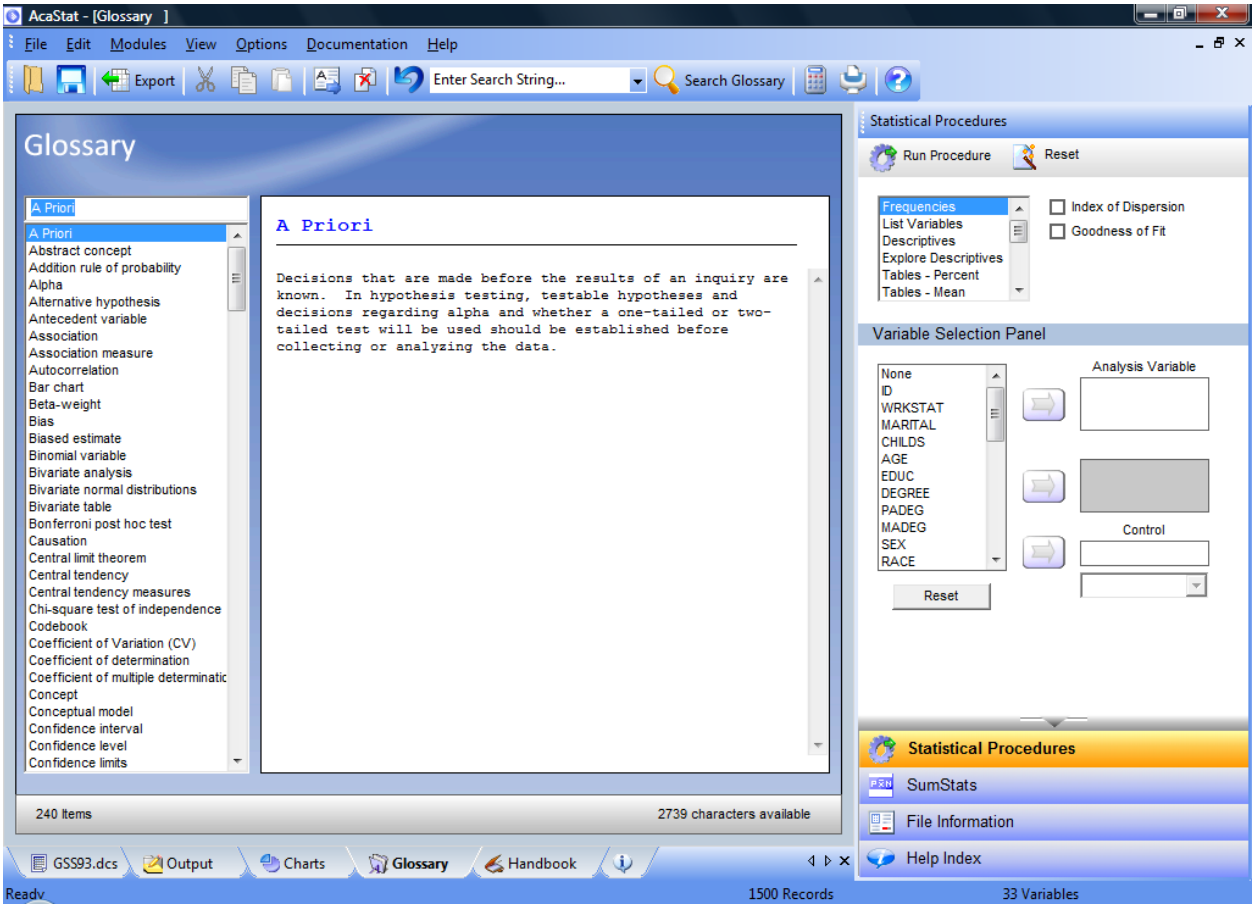
Glossary

Glossary provides a convenient location to store notes, definitions, and other reference information. Create and share your database with other Glossary users. You can also add the records from multiple files to create one list of notes.

AcaStat includes a statistical term glossary named "StatNotes.sn1". Edit or add new terms to the glossary. Create new Glossary files for recording project notes or storing output from analyses (copy from the Output Log). **IF YOU EDIT STATNOTES.SN1, YOU SHOULD SAVE IT UNDER A DIFFERENT NAME. DURING ROUTINE AUTOMATIC UPDATES AN UPDATED STATNOTES FILE MAY DOWNLOADED WHICH WILL REPLACE THE STATNOTES FILE ON YOUR COMPUTER.**

How Glossary Works

A list of the contents will be displayed to the left. Clicking an item in the list will display that record's contents. To edit the record, start typing in the relevant field. Changes are automatically saved when you click on another item. To create a new record, click the Add button on the toolbar. When done entering information, click Save on the toolbar.



Glossary File System

When you create a new file, Glossary establishes a file named "Temp" with the extension ".sn1" (e.g. "Temp.sn1"). Always save the Temp file under a different name before closing Glossary or creating another new file.

Searching Records

Glossary provides a search option for doing a keyword search of the active record set. Records with any match of the keyword (case insensitive) will be listed. Click on an item in the list to display the record. The items listed in the search can be edited or deleted. Improve the search potential by inserting common keywords at the end of a note. This will ensure that a search will find similar subjects.

Limitations

Glossary limits the number of characters per field. When the limit is reached, the field will stop allowing keyed entries. It is important to remember that spaces count as one character and paragraph returns count as two. The status bar at the bottom of the form will track the number of characters used.

The limitations of each field are as follows:

Subject : 60 characters
Note Text : 3000 characters (about one 8.5 x 11 page of text)

File Menu Items

New File

Creates a new file called "Temp.sn1". Save this file under a different name to keep the contents.

Open File

Opens an established Glossary file.

Save File As

Save a copy of the current file under a different name.

Add Records from Other Files

Select and add the contents of another Glossary file to the current file. This operation combines notes entered in other files into one convenient and searchable database.

Import

Use Import to create a new Glossary file that is populated with text created in other software. There are two options: a) paste the text into a new file or b) import a tab delimited text file. Both options work very well with spreadsheets.

Formatting is very important. The text must be contained in two columns. The first column should contain the subject and the second column should contain the descriptive text. Each row represents one record (subject). Once the text is formatted into the two columns, copy and paste the cell contents into Glossary using the Import option or save the file as a tab delimited text file for importing.

Export

Export will put the contents of the current Glossary file into a tab delimited text file. This file can be easily imported into other software. Exported files will have all paragraph returns removed from the descriptive text. As a result, if the descriptive text in Glossary was separated into several paragraphs, the resulting exported text file will combine these into one paragraph.

Print

Prints only the contents of the current record.

Edit Menu Items

Add Note

Creates a new record. Enter a subject label that will appear in the listing. Enter text in the space provided below the subject field.

Save Note

Saves the contents of the current record. By default, Glossary will sort the items in the list box and return to the first record.

Delete Note

Deletes the current note from the database.

Copy

Copies the currently selected text (shortcut [ctrl key + c]).

Paste

Pastes previously copied text into the field where your cursor is currently residing (shortcut [ctrl key + v]).

Search

Search records for a keyword or phrase. Records with a match will be displayed in the list box. Clicking on an item in the list box will display the record. The items listed in the search can be edited or deleted.

Reset Search

Click the Reset button to reveal the original list of records.

Practice Data Files

Data Grid data files are delimited text files but differ from import/export files in that they are configured to store variable labels, value labels, and missing values. There are several instructional data files provided with AcaStat. For any AcaStat formatted data file, a complete listing of variables is available by clicking the Data pull-down menu and selecting "Create Data Dictionary". Background information on the data, if previously entered and saved, can also be seen in the Project Notes tab.

Repair Practice Data Files

If a practice file becomes damaged or altered, replace the damaged file with an original version by clicking the Options pull-down menu and selecting "Repair Sample Data Files". If a practice data file has been deliberately altered, rename the file so this operation does not accidently replace the altered version.

Practice (csv) Delimited Data Files

Fictitious data are provided in one comma separated value (csv) file for practicing data importing.

File Name: "Income&Sex.cvs"
Location: AcaStat program folder, "\Data" subfolder

Fictitious Data (n=30)

VARIABLE NAME	DEFINITION
RECORD	Record number
INCOME	Annual income in US$
EDUC	Years of education
SEX	0=Male, 1=Female

To see what a comma or tab delimited text file looks like, open one of the csv or txt data files in the Output tab.

Manuals (pdf)

AcaStat has three manuals: User Manual, Student Workbook, and the Applied Statistics Handbook. They are electronic files in Adobe Acrobat Reader (pdf) format. To read these files, Adobe Acrobat Reader version 6.0 or later must be installed on the host computer. Acrobat Reader is a free download from Adobe. For more information, visit http://www.adobe.com for download information. Version 8 is included on the AcaStat CD purchase option.

Each manual includes an active contents page and can be viewed electronically or printed as reference material. The Student Workbook and the User Manual are available as free downloads from the AcaStat.com web site (www.acastat.com/pdfdocs.htm). If the manuals are installed in the application folder (C:\Program Files\AcaStat-6), they can be opened through the AcaStat Documentation menu. All AcaStat manuals are copyright protected.

User Manual

The User Manual contains instructions on how to use AcaStat. The content is similar to what is contained in AcaStat help but includes screenshots of AcaStat modules.

File Name: UserManual.pdf
Availability: AcaStat CD, Website http://www.acastat.com/pdfdocs.htm

Student Workbook

The Student Workbook contains over 20 lessons designed to introduce basic statistical concepts while also demonstrating the functionality of Data Grid, Stat Procs, and SumStats.

File Name: Workbook.pdf
Availability: AcaStat CD, Website http://www.acastat.com/pdfdocs.htm

Applied Statistics Handbook

The Applied Statistics Handbook is **not available in the Special Edition version** of AcaStat. It contains formulas, definitions, and examples, of both hand calculations and annotated output from analytical software. It is designed to be a quick reference resource (not a complete statistical text). The Applied Statistics Handbook can be opened through the Documentation pull-down menu in AcaStat. Purchase of the Licensed Edition of AcaStat is required. The Applied Statistics Handbook is a much expanded version of the html Research Methods Handbook included with the AcaStat program.

File Name: Statbook.pdf
Availability: Licensed Edition only of AcaStat.

Review of Features

AcaStat includes two data analysis modules (Data Grid, SumStats, and Charts) and three support modules containing analysis output (Output Log) and instructional material (Glossary and Handbook). Designed to provide students, teachers, researchers, and managers with an inexpensive and easy to use tool for basic data exploration. Special pricing and site license waivers are available for instructors adopting AcaStat for course work. Printable manuals include a Student Workbook with over 20 lessons, an Applied Statistics Handbook, and an AcaStat User Manual.

Data Grid

Analyze raw data with this spreadsheet design. Produces basic descriptive and inferential statistics, and creates crosstabulations and summary reports. Includes the features listed below. Export data directly to Excel. Run random sampling simulations.

Feature	Details
Spreadsheet	create data files, maximum edit capacity 200,000 data cells
Import Data	import delimited text files, paste tables from other software
Data formatting	create variable and value labels, set missing values
Variable Recode	compute, combine, and recode variables
Paste Data	copy and paste data from spreadsheets, documents, and database tables
Export Data	export comma and tab delimited value text files
Output Log	edit output, save as text file, print, copy, paste
Statistics	
Univariate summary statistics	frequencies (count, percent, total) mean, median, variance, standard deviation
Crosstabulations	count, row %, col %, total %
List observations	
Diagnostic accuracy	sensitivity, specificity, predictive values, and likelihood ratios
Summary reports	percentages, counts, means, sum, standard error, 95% CI
Two sample t-test of means	
One-way ANOVA	summary statistics and Bonferroni post hoc analysis
Chi-square	Cramer's V, Pearson C, Lambda, Kappa, Fisher's, odds ratio
Pearson correlation	correlation matrix
Simple regression	
Multiple regression	OLS and Logistic
Appraisal ratios	ratios, price-related differential, coefficient of dispersion, coefficient of variation
Graphing Features	histogram, scattergram, column, and pie charts (enlarge, print, copy)
Simulation	repeated random sampling to introduce the concept of sampling distributions.

SumStats

Analyze summary statistics (means, proportions, counts) to develop confidence intervals and conduct basic significance tests. Includes the features listed below. SumStats provides summary instructions and interpretations of output alongside the calculator.

Feature	Details
Statistical Tests	one and two sample t-test of means one-way ANOVA (4 sample) one and two sample z-test of proportions chi-square (5x5 table)
Margin of Error-Mean	90%-95%-99% confidence intervals, margin of error, upper limits, lower limits
Margin of Error-Proportions	90%-95%-99% confidence intervals, margin of error, upper limits, lower limits
Standardized Scores	z-score, graphic representation on normal distribution
Diagnostic Accuracy	sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio
Sample Estimator	power analysis for the difference between two proportions (z-tests) and two means (t-tests)

Output Log

Automatically stores all Data Grid and SumStats analyses for viewing, printing, and saving, and keeps a running record of all analyses. Print, save as text file, or use the one-click export to send directly to Microsoft Word.

Glossary

Note database for easy storage and retrieval. A statistics glossary of terms and definitions is provided. Edit and add new terms and definitions. Create your own database of terms or notes or store output from analyses.

Handbook

HTML statistics reference. Outlines hypothesis testing and provides formulas and examples of hand calculations for common statistical tests. Hyperlinked to SumStats modules for easy reference. Includes quick links to document webpage and links to websites providing reference material on statistics and research methods.

System Requirements (recommended minimum)

Windows95/98/NT/XP/2000/Vista
Pentium 700mhz/128 mb RAM
Download file size 15 MB; Installed size 30 MB
Install on host computer hard drive

AcaStat Software
43584 Merchant Mill Terrace
Leesburg, VA 20176 USA
support@acastat.com
http://www.acastat.com

Purchase Information

Prices include shipping and are subject to change. Most CD-ROM orders are shipped the next business day. Special orders, quantity discounts, and site licenses are available. Questions? Contact AcaStat Software via e-mail at info@acastat.com.

On line Orders (Credit Card)

Visit http://www.acastat.com for credit card orders, current prices, and site license purchases and renewals.

Mail orders (check, money-order, or purchase order/invoice)

AcaStat Software Mail Order Form

Name:_____ E-Mail: _____

Organization/School: _____

Street: _____ Phone: _____

City: _____ State/Province: _____

Country: _____ Postal Code: _____

Download - Retail $29.95 ea ____ x 29.95 = _____

Download – Academic $19.95 ea ____ x 19.95 = _____

Add Mini CD $10.00 ea ____ x 10.00 = _____

Subtotal = _____

Virginia residents add 5% sales tax (subtotal x .05) = _____

Total = _____

Make Check or Money Order Payable to:

AcaStat Software
43584 Merchant Mill Terrace
Leesburg, VA 20176 USA